

ESTIMATION AND PREDICTION OF HOSPITAL AND MEDICAL CARE COSTS

A project report submitted to Jawaharlal Nehru Technological University,

Kakinada

In the partial fulfillment for the award of the Degree of

BACHELOR OF TECHNOLOGY

IN

DATA SCIENCE

Submitted by

K. RISHITHA (20KE1A4420)

M. SOBHA REDDY (20KE1A4427)

M. HIMA BINDHU (20KE1A4424)

P. ASIFA KOUSAR (20KE1A4440)

S. VINDHYA VALI (20KE1A4452)

Under the Noble Guidance of



DEPARTMENT OF DATA SCIENCE

**MALINENI LAKAHMAIAH WOMEN'S ENGINEERING COLLEGE
(AUTONOMOUS)**

(An ISO 9001-2008 Certified & NBA Accredited Institution)

(Affiliated to Jawaharlal Nehru Technological University, Kakinada)

PULLADIGUNTA(VILLAGE), VATTICHERUKURU(MANDAL)

**MALINENI LAKAHMAIAH WOMEN'S ENGINEERING COLLEGE
(AUTONOMOUS)**

(An ISO 9001-2008 Certified & NBA Accredited Institution)
(Affiliated to Jawaharlal Nehru Technological University, Kakinada)
PULLADIGUNTA(VILLAGE), VATTICHERUKURU(MANDAL)



DEPARTMENT OF DATA SCIENCE

This is to certify that the project entitled “ESTIMATION AND PREDICTION OF HOSPITAL AND MEDICAL CARE COSTS” is a bonafide work of

K. Rishitha (20KE1A4420), M. Sobha Reddy (20KE1A4427), M. Hima Bindhu (20KE1A4424), P. Asifa Kousar (20KE1A4440), S. Vindhya Vali (20KE1A4452)

*in the partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in **DATA SCIENCE** and for the academic year **2023-2024**. This work is done under my supervision and guidance.*

Signature of the Guide

Yamini mam

Signature of the Head of the Department

Dr. HARI KRISHNA(HOD)

Signature of the External Examiner

ACKNOWLEDGMENT

“Task successfully” makes everyone happy. But the happiness will be gold without glitter if we didn't state the persons who have supported us to make it a success.

We would like to place on record the deep sense of gratitude to the Honorable chairperson Dr. Malineni Perumallu, Guntur for providing necessary facilities to carry the project work.

We express our gratitude to **Dr.J.APPARAO, Ph.D.**, principal of MLWEC, Guntur for his valuable suggestions and advices in the B.Tech course.

We express our gratitude to the **Head of the department of DS, Dr. HARI KRISHNA, MLWEC, Guntur** for his constant supervision, guidance and cooperation throughout the project.

We would like to express our thankfulness to our project guide **PRATHYUSHA MAM, MLWEC, Guntur** for his constant motivation and valuable help throughout the project work.

Finally we would like to thank our Parents, Family and friends for their co-operation to complete this project.

TEAM MEMBERS

K. Rishitha	(20KE1A4420)
M. Sobha Reddy	(20KE1A4427)
M. Hima Bindhu	(20KE1A4424)
P. Asifa Kousar	(20KE1A4440)
S. Vindhya Vali	(20KE1A4452)

DECLARATION

We hereby declare that the project work entitled “**ESTIMATION AND PREDICTION OF HOSPITAL AND MEDICAL CARE COSTS**” done under the guidance of **Mr. D. Ashok, Assistant professor**, is being submitted to the “Department of **DATA SCIENCE**”, **MALINENI LAKSHMAIAH WOMEN’S ENGINEERING COLLEGE**, Guntur is of our own and has not been submitted to any other university or Educational for any degree or diploma.

TEAM MEMBERS

K. Rishitha	(20KE1A4420)
M. Sobha Reddy	(20KE1A4427)
M. Hima Bindhu	(20KE1A4424)
P. Asifa Kousar	(20KE1A4440)
S. Vindhya Vali	(20KE1A4452)

TABLE OF CONTENTS

S/N	CHAPTER TITLE
1.	INTRODUCTION
2.	RELATED WORK
3.	THE PROPOSED METHOD BASED ON LINEAR REGRESSION
4.	RESULT AND ANALYSIS
5.	DISCUSSION
6.	CONCLUSION

ABSTRACT

Medical costs are one of the most common recurring expenses in a person's life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. The estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. Obesity prevention at a young age is a top concern in global health, clinical practice, and public health. To avoid these restrictions, genetic variants are employed as instrumental variables in this research. Using statistics from public huge datasets, the impact of body mass index (BMI) on overall healthcare expenses is predicted. A multi-view learning architecture can be used to leverage BMI information in records, including diagnostic texts, diagnostic IDs, and patient traits. A hierarchy perception structure was suggested to choose significant words, health checks, and diagnoses for training phase informative data representations, because various words, diagnoses, and previous health care have varying significance for expense calculation. In this system model, linear regression analysis, naive Bayes classifier, and random forest algorithms were compared using a business analytic method that applied statistical and machine learning approaches. According to the results of our forecasting method, linear regression has the maximum accuracy of 97.89 percent in forecasting overall healthcare costs. In terms of financial statistics, our methodology provides a predictive method.

1.INTRODUCTION:

Introduction to Business Intelligence:

- Business Intelligence
 1. Data Integration
 2. Data Processing
 3. Data Presentation
 4. ETL Architecture

Introduction to Data Analytics:

- Types of Data Analytics
 1. Descriptive Analytics
 2. Diagnostic Analytics
 3. Predictive Analytics
 4. Prescriptive Analytics

Analytics & Applications

Working with IBM Cognos Analytics:

- Introduction to Cognos Analytics
- Overview & Features
- Connecting Cognos Analytics to Data Sources
 1. Working with Flat files
 2. Connecting spreadsheets

Data Extraction:

- Introduction to Database
- Create IBM DB2 Service
- Creating Database & Table
- CRUD Operation on database tables

Cognos Analytics in Action:

- Exploratory Data Analysis EDA with Cognos Analytics
- Data Preprocessing

Present Data in Cognos Analytics:

- Creating Analytical Dashboards
- Generate personalized Reports & Story

Python for Data Science:

- Basics of Python
- Working with Data Structures
- Conditional Constructs
- Functions

Data Exploration & Visualization:

- Python Packages
 1. Working with NUMPY
 2. Working with Pandas
 3. Introduction to Data Visualization
 4. Matplotlib, Seaborn

Supervised Learning-Regression:

- Predictive Analytics with Regression
 1. Introduction to Regression Analysis
 2. Linear Regression Analysis
 3. Building predictive models using regression analysis

Supervised Learning-Classification:

- Predictive Analytics with Classification
 1. Introduction to Classification Analysis
 2. Building predictive models using classification analysis-Logistic Regression, Decision Tree & Random Forest

Build & Deploy ML Application:

- Introduction to different modes of development
- Working with Flask Frame-Work
- Building application with flask framework
- Integrating Machine Learning model with web application
- Embedding Dashboard, Report & Story with web application

Health care expenditure is one of the most critical issues in today's society. World Health Organization (WHO) statistics showed that global health care expenditure was approximately 7.5 trillion US\$, equivalent to 10% of the global GDP in 2016. One of the reasons for these high expenses in care are the low accountability in health care, such as unnecessary procedures or drugs used on patients, or excessive charges for patient treatments. If we could predict health care costs for each patient with high certainty, problems such as accountability could be solved, enabling control over all parties involved in patients' care. It could also be used for other applications such as risk assessment in the health insurance business, allowing competitive premium charges, or for the application of new policies by governments to improve public health. With the now-common use of electronic health records (EHR), an interest has emerged in solving accountability problems using data mining techniques. There have been various approaches to predict health care costs for large groups of people. On the contrary, prediction for an individual patient has rarely been tackled. Initially, rule-based methods were used for trying to solve these problems requiring domain knowledge as if-then rules. The downside of this method is the requirement of a domain expert to create the rules, thus making the solution expensive and limited to the dataset being used. In the current state of the art, statistical and supervised learning are preferred with supervised methods having better performance. The reason for better performance is the skewed and heavy right-hand tail with a spike at zero of the cost distribution in health care. Supervised learning methods can be evaluated by performance and interpretability; usually, the more sophisticated methods are the ones that have a better performance, sacrificing interpretation (e.g., Random Forest, Artificial Neural Networks, and Gradient Boosting). A drawback of these high performing machine learning algorithms in health care is their black-box nature, especially in critical use cases. Even though health care cost prediction is not a critical use case, using patients' personal and clinical information for this problem could suffer biased results without an interpretable method. Interpretable methods would allow patients, physicians, and insurers to understand the reasoning behind a prediction, giving them the option to accept or reject the knowledge the method is providing. In this work, we present an interpretable regression method applied to the health care cost prediction problem based on the Dempster-Shafer theory, also known as the theory of belief function. We based our work on Petit-Renaud and Denoeux evidence regression model using a discount function related to the importance of each dimension. Each dimension importance is learned during the training phase in two different approaches. The first approach uses a variable for each dimension, and the other one uses an Artificial Neural Network (ANN) to obtain the weights of the dimensions. In both approaches, the optimal parameters are learned using gradient descent. Given the transparency of the evidence regression model, we create a set of rules for each patient in the training set based on their

vicinity, and when a prediction is made, we give the set of rules with their importance. Our research question is whether it is possible to develop an interpretable method that has a performance similar to black-box methods for the health care cost prediction problem. To test our answer, we used Japanese health records from Tsuyama Chuo Hospital, which include medical checkups, exam results, and billing information from 2013 to 2018, and compare our method performance with less interpretable methods such as Random Forest, ANN, and Gradient boosting (GB). Our results show that our transparent model outperforms the ANN and GB models in the health care cost predict with an R^2 of 1.44.

The incidence of overweight and obesity has increased significantly in most countries in recent decades. Excess weight is associated with an increased incidence of many chronic diseases, including vascular disease, respiratory disease, osteoarthritis, some cancer, type 2 diabetes, and premature death. There is consistent evidence that an increased BMI is associated with higher health costs, and these costs are expected to increase as obesity. Modelling uses machine-learning methods, in which the machine learns from the data and uses it to forecast new data.

The most commonly predictive analytic model used is regression. The proposed model for accurate prediction of future outputs has applications in banking, economics, e-commerce, sports, business, entertainment, etc. A method used to forecast healthcare costs for BMI is based on several factors. Multiple linear regression is one of the statistical techniques for estimating the relationship among the dependent (target) and independent variables. The regression method is commonly used to develop a system based on a number of factors to predict the cost.

The regression analysis is performed to determine the relationship among two or more variables with cause-effect relationships and to make predictions for the topic using the relationships. If regression used one independent variable, then it is known as univariate regression analysis, or else if it used more than two independent variables then it is known as multivariate regression analysis. Linear regression involves initially uploading the data and then analysing the data. Subsequently, the data are cut, and then, the data are trained and separated to create the model. At last, it will evaluate the accuracy. The main aim of regression is to develop an efficient technique for predicting dependent properties from a set of characteristic variables. A regression problem is the actual or continuous value of the output variables, that is, area, salary, and weight. Regression can be defined as a statistical method used in applications such as predicting the healthcare costs. Regression is used to predict the relationship among the dependent variable and set of independent variables. There are various types of regression techniques available namely simple linear regression, multiple linear regression, polynomial regression, support vector regression, and random forest regression.

Fast-growing healthcare costs have become a significant challenge in several developed countries. Existing evidence suggests that healthcare costs have accumulated among a large number of BMI. Even though experiments have

attempted to develop accurate models for predicting healthcare costs for BMI, their effectiveness is excellent due to the lack of detailed clinical information in the data used to create complex intervals and prognostic models. Numerous studies on more costs for obesity patient prognostic models have relied on self-report data and electronic health data from claims [14]. Data from laboratory tests are defined—these, more granular and detailed clinical information, lead to improvements in the prognostic model. A recent survey by health research program and claim data shows that there is an improvement in the performance of the machine learning-based predictive model for health costs for obesity. Still, many insurers and providers worldwide are actively seeking an approach that can accurately predict obesity BMI. However, despite the potential value of advanced machine-learning approaches for risk prediction, payers and providers still rely heavily on linear regression to manage and adapt their patient population. The slow adoption of advanced machine-learning techniques may be partly explained by the lack of familiarity with risk stabilization analysts with such techniques and the combination of complex interpretation and results required in practice. Machine-learning regression models are within the framework of standard linear regression and perform some sophisticated but less explicit machine-learning techniques. This study focused on fine linear regression models, which conducted a complete comparison of penalty regression with linear regression in forecasting overall health costs, which was not reported in the previously published literature. The major focus of this study is to estimate the health costs incurred due to obesity in the population.

The rest of this study is formalized as follows: Section 2 defines the related works on estimating the healthcare costs using various methodology methods. Section 3 designates in detail the workflow of the proposed algorithm. Section 4 represents the experiments with results and comparison graphics with existing works and its discussion. Finally, Section 6 concludes the study.

2.RELATED WORK:

Some of the recent literature that describes the various mechanism of estimating the costs of physical healthcare is summarized below. In, unplanned 30-day readmissions are a common occurrence among congestive heart failure (CHF) patients, posing major health concerns and increasing healthcare costs. It is critical to implement tailored treatment programs for high hazard patients of readmission in an attempt to prevent readmissions and lower healthcare costs. This necessitates recognizing high individuals at the time of hospital release. They constructed and evaluated a deep learning network to predict 30-day unplanned readmission using actual information from over 7,500 CHF patients hospitalized in Sweden. Using specialist characteristics and situational integration of medical knowledge provides a cost-sensitive implementation of the long short-term memory (LSTM) neural net. Using both machine derived and professional characteristics, including frequent patterns, and resolving the issue of class imbalances, this research focuses on important parts of an EHR-driven forecasting system in a single framework. We assess each element's impact on forecasting effectiveness (F1 measure, ROC-AUC) and price benefits. In at least 2 evaluating criteria, it shows that the technique with all critical features outperforms the simplified approaches in terms of discriminating capability. Researchers also propose a basic economic assessment to predict annual income if high-risk patients are provided tailored therapies.

Patients with heart failure (HF) require precise hazard classification to implement tailored therapies focused on enhancing their efficiency of living and results. To assess the economic benefit of complementing claim-based forecasting analytics with electronic medical record (EMR)derived data and to contrast machine-learning techniques to conventional logistic regression in forecasting critical results in patients with HF, healthcare patients with HF from 2 healthcare professional systems in Massachusetts, Boston, were included in predictive research with a one-year follows up duration. "Providers" comprise therapists, various medical professionals, clinicians, and their organization including the network. Logistic regression, gradient boosted modelling, regression trees, random forests, least absolute shrinkage, classification, and selection operation regression were used to predict all-cause morbidity, top cost decile, HF hospitalization, gradient boosted modelling, and home days loss larger than 25%. Information from network 1 was used to educate all algorithms, which were then evaluated in network 2. The area under high accuracy curves (AUPRCs) and overall value estimations from decision curves were obtained after choosing the best effective modelling strategy depending on the Brier score, calibration, and discrimination.

The goal of this study was to evaluate the effectiveness of machine-learning methodologies for predicting healthcare expenses connected with spinal fusion in aspects of gains or losses in Taiwan Diagnosis-Related Groups (Tw-DRGs) and to use these techniques to investigate the major features connected with spinal fusion medical costs. Methods: a data collection was gathered from a healthcare facility

centre in Taoyuan, Taiwan, containing data on Tw-DRG49702 patients (without problems or comorbidity; posterior and other spinal fusion). Weka 3.8.1 was used to forecast using random forest, support vector machines, Naive Bayesian, C4.5 decision tree, and logistic regression approaches. The research showed that the random forest approach may be used to estimate the healthcare expenditures of Tw-DRG49702 and that it can help institutions improve the financially operational effectiveness of this procedure.

Because of the ageing populations and enhanced therapy of fundamental conditions, cardiac arrest is among the most complicated chronic disorders with a higher incidence. The incidence is projected to gradually climb, reaching 3% of the population in Western countries. It is the leading reason for hospitalizations in people aged 65 and above, leading to substantial expenses and a significant societal effect. In the therapy of HF, the present “one-size-fits-all” strategy does not produce the optimal results for all patients. These facts pose a serious danger to the proper treatment of heart failure patients. It will take an unconventional method from a unique perspective on health care. We offer a unique forecasting, preventive, and personalized healthcare strategy, in which patients are actually in charge of their care, aided by a user-friendly online form that employs artificial intelligence (AI). This technique study outlines the demands in HF care, as well as the necessary paradigm shift and the factors necessary to make it happen. A digital physician is being developed through an exciting combination of medical and high-tech partners from patient coaching, serious gaming, North-West Europe, artificial intelligence, and combining state-of-the-art HF health care. The findings are intended to improve and customize self-care, in which patients conduct routine care chores without the intervention of healthcare experts, allowing them to focus on more difficult problems. This innovative approach to health care will lower prices per patient while increasing results, ensuring the long-term viability of top-tier HF health care.

In, DRG codes are useful for price tracking and allocation of resources since healthcare operators obtain predetermined levels of compensation for certain treatments under diagnosis-related group (DRG) payments. Coding, on the other hand, is usually done after the fact, after the patient has been discharged. They want to use normal medical text to forecast DRGs and DRG-based case mix index (CMI) at initial inpatient admission to forecast hospital costs in an acute context. Without manual coding, a deep learning based natural language processing (NLP) method is tested to forecast cost-reflecting weights and per-episode DRGs on 2 cohorts (paid by All Patient Refined (APR) DRG or Medicare Severity (MS) DRG). In fivefold cross-validation trials on the first day of ICU admission, it attained macro averaged area under the receiver operating characteristic curve (AUC) scores of 0.871 (SD 0.011) on MS-DRG and 0.884 (0.003) on APR-DRG. When applied to hypothetical patient populations to predict average cost-reflecting weights, the algorithm improved over time, yielding absolute CMI errors of 12.79 (2.31%) and 2.40 (1.07%) on the first day, correspondingly. Because the system can adjust to changes in admission time and cohort size while requiring no additional manual coding, it

has the potential to aid in cost estimation for active patients and enable improved functional outcome in hospitals.

2.1. Health Care Cost Prediction:

Statistical methods (e.g., linear regression) suffer from the spike at zero and skewed distribution with a heavy right-hand tail of health care costs in small to medium sample sizes. Advanced methods have been proposed to address this problem, for example, Generalized Linear Models (GLM) where a mean function (between the linear predictor and the mean) and a variance function (between the mean and variance on the original scale) are specified and the parameters are estimated given these structural assumptions. Another example is the two-part and hurdle model, where a Logit or Probit model is used in the first instance to estimate the probability of the cost been zero, and then if it is not, a statistical model is applied, such as Log-linear or GLM. The most complex statistical method used to solve this problem is the Markov Chain model; an approach based on a finite Markov chain suggested estimating resource use over different phases of health care. Mihaylova et al. present a detailed comparison of statistical methods in health care cost prediction. Supervised learning methods have been vastly used to predict health care costs; the data used for these methods vary. While a few works use only demographic and clinical information (e.g., diagnosis groups, number of admissions and number of laboratory tests), the majority have incorporated cost inputs (e.g., previous total costs, previous medication costs) as well, obtaining better performance. GB excels as the method with the best performance for this problem, which is an ensemble-learning algorithm, where the final model is an ensemble of weak regression tree models, which are built in a forward stage-wise fashion. The most essential attribute of the algorithm is that it combines the models by allowing optimization of an arbitrary loss function, in other words, each Proceedings 2019, 31, 74 3 of 13 regression tree is fitted on the negative gradient of the given loss function, which is set to the least absolute deviation. ANNs come close to the performance of GB, ANNs are an extensive collection of processing units (i.e., neurons), where each unit is connected with many others; ANNs typically consist of multiple layers, and some goal is to solve problems in the same way that the human brain would do it. Another type of model with good results is the M5 Tree; this algorithm is also a Regression Tree, where a Linear Regression Model is used for building the model and calculating the sum of errors as opposed to the mean. In health care, the majority of the expenses of a population are originated from a small group, as Bertsimas et al. Showed in their dataset: 80% of the overall cost of the population originates from only 20% of the most expensive members. Therefore, to improve the performance of the methods listed above, a classification phase is suggested to classify patients in a risk bucket. Morid et al. Reported that for low-risk buckets, GB obtains the best results, but for higher ones, ANN is recommended.

2.2. Interpretability:

In machine learning, interpretability is the ability of a model to explain or present its prediction in an understandable way. The techniques used for interpretability fall into two categories. The first is model transparency, which means to fully comprehend a model, understanding the parts of the model, input values, parameters, and calculation; there is an intuitive explanation, and it can prove that training will converge to a unique solution. The second category is post-hoc explanations, usually applied to black-box models, where a prediction is presented in a comprehensible way with visual or textual artifacts. In some domains (e.g., medical domain), method interpretability can be as important as its accuracy or even more, given legal or ethical reasons. Interpretability also helps to gain confidence from its end-users; this is why for their ease of interpretation, some problems use simple, transparent models with less accuracy instead of complex, more accurate ones. For example, Cuarana et al. used generalized additive models with pairwise interactions applied to predict pneumonia risk and hospital 30-day readmission, Ustun et al. Created a data-driven scoring system called a Super-sparse Linear Integer Model to create a highly tailored scoring system for sleep apnea screening. Naive-Bayes has been used to create a prediction system for heart disease. Regression and decision trees have been applied to a variety of problems. Some authors have used the Dempster-Shafer theory, also called the theory of belief functions, which is a generalization of the Bayesian theory, like Maseleno et al., who created an expert system for the detection of skin diseases and Penafiel et al. who associated the risk of getting a stroke with health checkup data. Today with the heavy adoption of black-box methods (e.g., ANN), there has been a need to interpret these models to improve their performance and make them more reliable. The complex relations between its attributes make them unintelligible, but they usually outperform transparent models. Some of these approaches are focused exclusively on interpreting ANN models. Others treat the models as black-box functions developing model-agnostic approaches that produce post-hoc explanations. One of these approaches is to use transparent models to create an approximate representation of the black-box method. Another approach is using perturbation-based methods, which consists of making perturbations in individual inputs and observe the variation of the output. This approach has similarities with a sensitivity analysis process of a model. Ribeiro et al. used these two approaches and created LIME, an algorithm that can explain the prediction presenting textual or visual artifacts that provide a qualitative understanding of their relationship between the components of the instance and the model prediction of any classifier or regressor by approximating it locally with an interpretable method.

2.3. Dempster Shafer Theory

The Dempster-Shafer Theory (DST) is a generalization of the Bayesian theory that is more expressive than classical Bayesian models since it allows to assign “masses” to multiple outcomes measuring the degree of uncertainty of the process. Let X be the set of all states of a system called frame of discernment. A mass assignment function m is a function that satisfies:

$$m : 2^X \rightarrow [0, 1], \quad m(\emptyset) = 0, \quad \sum_{A \subseteq X} m(A) = 1$$

The term $m(A)$ can be interpreted as the probability of getting precisely the outcome A , and not a subset of A . Multiple evidence sources expressed by their mass assignment functions of the same frame of discernment can be combined using the Dempster Rule (DR). Given two mass assignment functions m_1 and m_2 , a new mass assignment function m_c can be constructed by the combination of the other two using the following formula:

$$m_c(A) = m_1(A) \oplus m_2(A) = \frac{1}{K} \sum_{B \cap C = A, B \neq \emptyset, C \neq \emptyset} m_1(B) m_2(C)$$

where K is a constant representing the degree of conflict between m_1 and m_2 and is given by the following expression:

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C).$$

Petit-Renaud and Denoeux were the first ones to introduce a regression analysis based on a fuzzy extension of belief functions, called evidence regression (EVREG). Given an input vector, they predict a target variable in the form of a collection of evidence associated with a mass of belief. This evidence can be fuzzy sets, numbers, or intervals, which are obtained from a training set based on a discounting function that takes their distance to the input vector and is pooled using the Dempster combination rule. They showed that their methods work better than similar standard regression techniques such as the Nearest Neighbours using data of a simulated impact of motorcycle with an obstacle. The EVREG model has been used for predicting the time at which a system or a component will no longer perform its intended function (machinery prognostic) for industrial application. Niu and Yang used the EVREG model to construct time series, whose prediction results are validated using condition monitoring data of a methane compressor to predict the degradation trend. They compared the results of the EVREG model with six statistical indexes, resulting in a better performance of the EVREG model. Baraldi et al. used the model to estimate the remaining useful life of the equipment. Their results have shown the effectiveness of the EVREG method for uncertainty

treatment and its superiority over the Kernel Density Estimation and the Mean-Variance Estimation methods in terms of reliability and precision.

3.THE PROPOSED METHOD BASED ON LINEAR REGRESSION:

Linear regression is one of the most common supervisory Machine learning statistical analysis techniques. It is commonly used to find linear correlations between two or more responses and predictive variables. The technique is divided into two types depending on the number of variables in the model such as simple linear regression and multiple linear regression. A response variable corresponding to a predictive variable is simple linear regression. Whether more than two response variables correspond to predictive variables is known as multiple linear regression as shown in Figure 1. This work used linear regression to study the relationship among total maintenance and other properties in datasets to obtain the properties most affected by the total cost of maintenance. 75% of the data in the dataset were trained, and 25% of the data were tested. Then, Pearson's correlation coefficient (PCC) for each simple linear regression sample was calculated. The PCC is determined and calculated by the following equation to find the parallel variability and strength of a linear regression relationship between two factors:

$$Y'_i = f_n(X'_i, \theta_p) + e$$

Here, X'_i and y'_i represent the independent variable and dependent variable; f_n represents the function; θ_p represents the unknown parameters; and e represents the error terms. The most commonly used measurements to estimate the performance of a linear regression are the root mean square error (RMSE), the mean absolute error (MAE), and the mean square error (MSE). The following equations denote the error deviation for regression:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}}$$

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

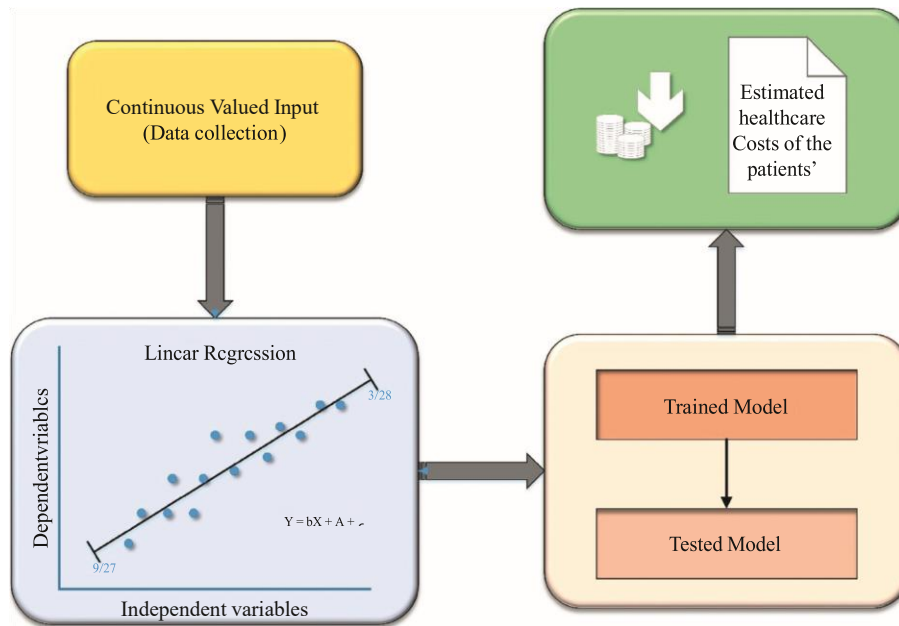


Figure:1 Block diagram for the proposal model

3.1 Regression's Role in predicting the Costs:

Clinics are encouraged to find more meaning in the substantial amount of data they generate and store each day. Regression provides useful predictive accuracy and value for machine learning clinics' databases with useful methods, features, and structures and contributes to a variety of strategies. The regression method aims to identify the possibility of improving results based on the predictive value of large-scale datasets for annual health costs. This is evidence of effectiveness in dealing with priority tasks, which defines that behaviours have the maximum tendency to cause preferred outcomes.

3.2. Steps for Applying Regression to Datasets:

The database used here is a collection of medical expense personal data, which contain anonymous information about people. These data will act as a method learning object to generate functional information. In Table 1, the attributes such as BMI and age are continuous variables, and the attributes such as smoker and sex are categorical variables:

- The next step is data exploration and preparation, and the quality of any machine-learning program is largely based on the quality of the data it uses. This stage requires more human intervention in the machine-learning process. Frequently cited statistics show that 80% of efforts in machine learning are dedicated to data. Most of this time is spent learning more about data and its nuances throughout an exercise known as data analysis.
- Then, a model on the data is trained. The specific machine-learning task will announce the selection of the suitable method, and the method will denote the data in the form of a model.

Attributes	Specifications
BMI	Body mass index
Age	Primary beneficiary age
Sex	Gender (male/female)
Smoker	The one who smokes affected by the obesity
Children	Number of children under BMI
Costs	Individual healthcare costs of the respective person

- Subsequently, the model performance is evaluated. It is important to evaluate how well the method has learned from its past experience as each machine learning model results in a biased solution to the learning problem. Depending on the type of model used, the accuracy of the sample can be estimated using the experimental database.
- Finally, the performance of the model is improved. It's necessary to use advanced techniques to increase the performance of the model if better performance is required. Each time, an entirely different type of model may have to be changed. After completing these steps, if the model appears to be operating acceptably, it can be used for its intended purpose. This model can be used to provide score data for forecasting, for financial data forecasting, to generate relevant insights for marketing or research, or to automate tasks.

3.3. Dataset Description:

We intended to forecast a patient's healthcare costs for the coming year depending on their insurance payment statistics and previous healthcare data. Tsuyama Chuo Hospital contributed the healthcare record information. These documents come from healthcare insurance applications that the hospital is required to submit to the administration. Every patient is recognized by an individual identity (ID) in these reports, which include the patient's conditions, medications, operations, and payment details. This claim's comprehensive paperwork can be obtained on the relevant website. We were able to retrieve the following information using this information:

- (i) Patient demographics include age and gender.
- (ii) Patients' characteristics include their body fat percentage, height, weight, and waist circumference.
- (iii) Health care verifies the outcomes of a patient's healthcare check-up tests. Every testing is assigned a code, and the outcome should be provided. Blood pressure (BP) and creatinine levels are two instances. There are 25 various categories of tests, as well as the date that they were gathered.
- (iv) Prognosis: a patient's ailment is diagnosed using ICD-10 codes and is tracked by date.
- (v) Payment details: for every session or hospital stay, every patient was assigned a score. This result effectively corresponds to the expense of a patient payment, which is the figure we needed to forecast for the following years.

It has been demonstrated that predicting patients' healthcare costs solely based on medical data is difficult. Preceding healthcare expenses are the strongest predictor of future expenditures: a longer history of healthcare expenditures is considered to increase forecasting. Depending on this fact, it is easier to anticipate future healthcare expenses when patients' information is available for multiple periods. When attempting to forecast expenditures for a single year, at least a two-year history is required.

Patients' monthly histories were included in our database. Furthermore, since many patients only had limited claims per year, there are several missing data. As a result, we decided to arrange claims by year to reduce the number of missing information. This technique did not work out as planned because many patients only had data. We next screened out these patients, leaving only those with clinical history. The fundamental characteristics of these patients are shown in Table 2.

Figure 2 forecasts every patient's scores for the following year. These scores are directly proportional to the amount of cost a patient spent on health care. The range of patient values is depicted in the graph. As anticipated for healthcare expenses, the scores exhibit all similar patterns as indicated previously, with a spike at zero and a lengthy right-hand tail.

It has been claimed that using medical characteristics produces similar results as using solely expense predictions. Although the fact that medical record appears to have little effect on forecast accuracy, we choose to maintain it because it can enhance the range of variables in the algorithm, which might enhance vector differentiation. Every resource accessible as characteristics was used to encode a patient's history. Demographics, health check-up results, ICD-10 diagnostic groupings, real score, and preceding score are the inputs. The patient's vector is described in full in the table. We employed all of the parameters listed in the table as input vectors, with the exception of the real score, which was used as our target attribute.

Table 2: Patients' characteristics and their predicted value.

Statistics	Predicted value
Total no. of patients	24,353
Mean value for expenses	10,538
Mean (age)	46.08
Male (%)	47.48
Female (%)	50.30

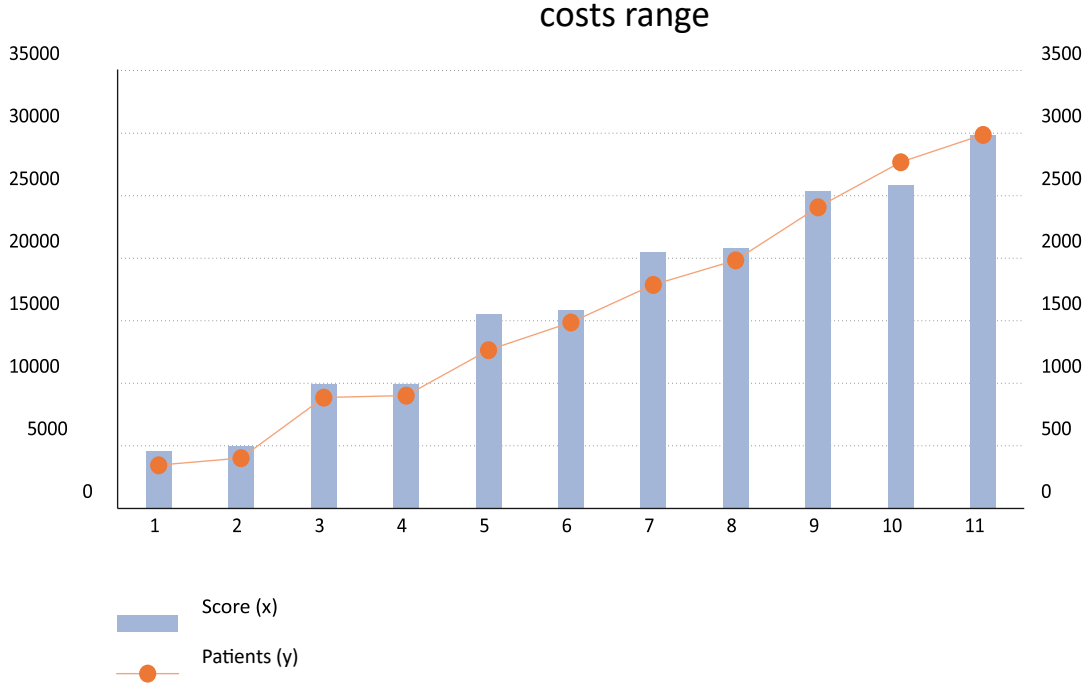


Figure 2: Graphic representation of cost range for patients' score.

3.4. Training Phase:

We must determine the ideal hyperparameters of our system for a forecast to adapt as closely as feasible to its true value. The weights of every dimension used in the distance function and g in the discount function are these parameters. For the training process, we used the gradient-based methods since they have a strong mathematical foundation for achieving optimal results.

The gradient descent technique is an automated approach for minimizing or maximizing a target function by optimizing variable values. As our objective parameter, we used the mean absolute error (MAE), which is calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

The target is to minimize the values of the MAE equation, which is dependent on the variable v_t that could be either c or ω . The following equation gives the updated value of v_t , termed v_{t+1} as follows:

$$v_{t+1} = v_t - \alpha \frac{zMAE}{zV}$$

This technique offers us a series of numbers for v_0, \dots, v_k that minimizes the MAE, with the first value for v (*i.e.*, v_0) generally chosen at random. During the training phase, we use all of the remaining $N - 1$ patients in L as evidence to try to forecast the expense of every patient p_i in the training set L . An epoch is an execution that computes forecasts for every ' N ' patient; the gradient descent approach accumulates by completing repeated epochs.

3.5. Time Optimizing in Computing:

A prediction's computing duration scales linearly with the size of the training phase. To find the mass of vectors of dimension m in a database with a training dataset of size n , we must firstly use the discounting function, which has a $O(m)$ complexity. With the training set, we can estimate any discounting functions of the input vector in $O(mn)$. Then, we can estimate K (9), which requires $O(mn)$. for every output series and $O(mn)$ for the accumulation; thus, we can estimate K in $O(mn)$. time. Lastly, we require the discounting function, K , and a product series to get the mass, so we estimate the weights of the input vector (8) while keeping $O(mn)$. Therefore, given $O(mn)$. complexity, we could obtain the forecast.

According to reference, a K -nearest neighbour technique could be used to accelerate up calculation without sacrificing efficiency. For the actual closest neighbour's searches depending on product quantization, we used methodology. Using this technique, we can generate indices for the K -nearest searches in time $O(mn + Kn)$ within the training step. The weights of the K -nearest neighbours, which will be estimated in $O(Km)$, are thus all that is required for a fresh forecast; the other weights are presumed to be null. Whenever the algorithm has been trained, the prediction's complexity is $O(Km)$.

3.6. Interpretability:

The IEVREG is a framework that is accessible. For every forecasting we generate, we could calculate the proportion (mass) of every element of information in the testing phase L . As a result, we have a complete understanding of how the anticipated quantity is calculated. This prototype is already interpretable, but to make it completely understandable, we will write a system of regulations for every forecast using the weights from the training dataset and the masses of every dimension gained all

through the training step. The idea is to calculate how much every piece of proof adds to the forecast. Firstly, using the weights of the existing $N1$ patients in the training dataset, we establish a system of regulations for each of the patients in the training phase for forecasting. Using the weights of the remaining $N1$ patients in the

phases and the weights of the dimensions, we firstly build a system of regulations for each of the patients in the training phase. The limits of the measurements for each of the input characteristics, as well as their weights, are encoded by these principles. The algorithm then chooses the patients in the training phase who are the most identical and combines their principles to generate a new collection of criteria for that forecast.

We use a tiny healthcare coverage database only with 5 characteristics as input to demonstrate how we get the regulations with the IEVREG framework. Table 3 shows the 5 data inputs (measurements) and the anticipated result for the healthcare expenses.

We used only the 60 closest neighbours to forecast this patient's result. The most significant principles (greater values) for expense forecasting are then obtained, as illustrated in Table 4. These are the limits and parameters that the patients have in common with the patients in the training phase.

We could see how a patient's expense projection is interpreted in Table 4. Low weight is associated with age in the IEVREG framework, while higher weight is associated with others. As a result, the method seeks out individuals with identical genders, BMIs, children, and smoking statuses, while ignoring age.

Algorithm 1 represents the steps of the linear regression model.

The flowchart for the proposed linear regression model is shown in Figure 3.

- Data And Problem Description:

We wanted to predict the health care cost of a patient in the future year based on their past medical records and health insurance billing information. The health records data was provided by the Tsuyama Chuo Hospital, from 2016 and 2017. These records are obtained from health insurance claims that the hospital must report to the Japanese Government. In these claims, each patient can be identified by a unique id and contains the patient's information of symptoms, treatments, procedures, and billing. The detailed documentation of this claim can be found at

http://www.iryohoken.go.jp/shinryohoshu/file/spec/22bt1_1_kiroku.pdf. We used this data to obtain:

Demographics: Patient gender and age.

Patient's attributes: General information about patients such as height, weight, body fat, and waist measurement.

Health checks: Results from health check exams a patient had undergone. Japanese workers undergo these exams annually by law. A code indexes each exam, and the result is also included. Some examples are creatinine levels and blood pressure. There are 28 different types of exams, and the date when they were collected is also included.

Diagnosis: Diagnosis for a patient illness registered by date and identified by their ICD-10 codes.

Billing information: Each patient had a score registered for each visit or stay in the hospital. This score translates directly to the cost of a patient bill and this is the value we wanted to predict for the next year. As shown by, it is challenging to predict patients' health care costs by only using clinical information. The best indicator for future health care costs are previous costs: the additional history of health care expenses known to improve the prediction. Based on this fact, prediction of future health care costs is better done when patients' data is known for consecutive periods. At least 2 years history is needed when trying to predict the costs for one year. Our dataset had patients' monthly history for 2016 and 2017. However, there are many missing values because most patients had few claims each year. Therefore, we chose to group claims yearly so that we could have fewer missing values. This strategy did not work as expected since many patients had data only for 2016. We then filtered these patients out, and thus, the final set of patients are those who have clinical history for both 2016 and 2017. Table 1 shows the basic statistics of this patients.

Table 1. Statistics of patients' records

Statistics	values
Total number of patients	25,464
Mean score for costs	10,649
Mean age	47.09
% Male	48.59
% Female	51.41

The value we needed to predict is the score of each patient in the future year. This score translates directly to the money a patient paid for healthcare. Figure 1 shows the distribution of patients' scores. The chart shows that the score has the same distribution as described in, with a spike at 0 and a long right-hand tail as expected for health care cost.

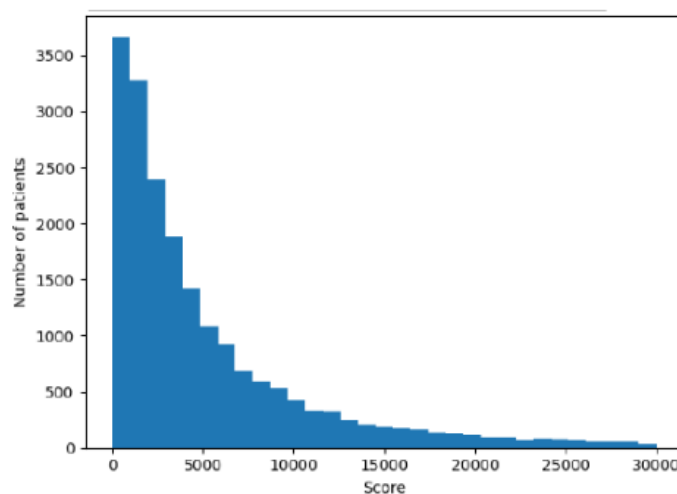


Figure 1. Patients score distribution

It has been reported that the use of clinical features yields the same performance as using only cost predictors. Despite clinical information seems not to affect prediction performance, we prefer to keep it, because having it in the model could increase the number of dimensions which may improve vector differentiation. Encoding a patient's history was done by using all sources available as features. The sources are demographics, health checkups result, ICD-10 diagnosis groups, previous score, and actual score. Table2 shows a detailed description of the patient's vector. As input vector, we used all dimensions shown in Table2 except for the actual score that was used as our target variable

Table 2. Patients encoding

Variable Number	Description
1-2	Demographics
3-30	Health checkup results
31-51	ICD-10 Diagnosis groups
52	Previous score
53	Actual score

4.RESULTS AND ANALYSIS:

The average annual rates and costs of consultations, tests, and prescription items were estimated by BMI category at the time of recruitment as shown in Figure 4. Percentage differences in rates and average annual costs were calculated for women with a BMI greater than 2 kg/m² and a BMI greater than 20 kg/m², both overall and according to the type of drug use. All models were evaluated using semi-possible generalized linear models with variations such as record link and Poisson. At the beginning of each year, annual expenses are estimated in subgroups defined by alcohol consumption, socioeconomic status, smoking level, educational qualifications, and strenuous exercise in recruitment [37]. The diversity of the proportional increases in annual costs among the types of each subgroup was estimated using the chisquare test.

The mean absolute error, moreover, is ineffective for comparing outcomes with costs stated in various dollars, so we will use the mean absolute percentage error (MAPE), a customized absolute error in which the MAE is reduced by the mean cost and calculated as follow

$$\text{MAPE} = \frac{(1/n) \sum_{i=1}^n ||y_i - \hat{y}_i||}{m}.$$

Here, \hat{y}_i is the estimated output for parameter y_i and m is the mean of variable y , denoted as follows:

$$m = \frac{1}{n} \sum_{i=1}^n y_i.$$

We will also use additional metric, the R^2 , which reflects how closely we are to the true cost curve and is defined as the Pearson correlation among projected and real healthcare costs. The following formula is used to determine this significance:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - m)^2}.$$

- Proposed model:

4.1Model implementation:

In this work, we extended the evidence regression model (EVREG) proposed by Petit-Renaud and Denoeux to be applied to the prediction of health care cost; we called this method the Interpretable Evidence Regression (IEVREG) model.

To predict a patient p_i health care cost (y), we use a set of other patients as evidence. First, we compute a mass (m_i) of each patient in the evidence set; this mass represents the similarity of the evidence patients with p_i (the one for whom we want to predict the health care cost). Then the target variable y (health care cost) can be calculated as the expected value of the mass m_i and target value y_i of each evidence patient.

$$E[y] = \sum_{i=1}^N m_i * y_i$$

formally, we define the training set as:

$$\mathcal{L} : \{p_i = (x_i, y_i)\}_{i=1}^N$$

Where x_i is the input vector of patient p_i and y_i the actual score (health care cost) for this patient (our target variable). To compute this cost, we need to obtain the target value y_i the training set L . Each patient in L is a potential evidence to discover the value y_i , all patients in the training set have a mass (m_i) that represents their similarity with the patient which cost we are trying to predict. The training set L has also an upper and lower bound for the y variables (max and min-cost), so besides each patient in the training set, the domain of the variable y is also considered another piece of evidence. The similarity of the patient whom we are predicting the costs, with the ones in L is measured by a distance function, $\mathbb{R}^n \rightarrow \mathbb{R}$ and a discount function $\phi(d(x_i, x_j))$, $\mathbb{R} \rightarrow [0, 1]$ which takes the input vectors of the patients. When the distance between the vectors is 0, the discount function is 1, and when the distance is infinite, the discount function is 0. We defined the distance d as

$$d(x_i, x_j) = \|(x_i - x_j) * w\|$$

Where w is a vector of the same dimension as x_i and x_j , representing the weights for each dimension, i.e., the amount which a dimension (age, gender, the result of an exam) contributes to the distance between two patients, thus the weight is the importance of each feature. For this purpose, the values of the input vector should be normalized (e.g., all values must be between 0 and 1).

The discount function we used is defined as:

$$\phi(d) = e^{-\frac{d^2}{\gamma}} \quad (7)$$

where w and γ are values learned during the training phase, in our case we started with $\gamma = 1$ and w as a vector containing only ones. Then the mass of each patient in \mathcal{L} is computed using Dempster rule of combination (2) obtaining:

$$m_j(x_i) = \frac{1}{K} \phi(d(x_i, x_j)) \prod_{h \neq j} (1 - \phi(d(x_i, x_h))) \quad (8)$$

where,

$$K = \prod_{j=1}^N (1 - \phi(d(x_i, x_j))) + \sum_{j=1}^N \phi(d(x_i, x_j)) \prod_{h \neq j} (1 - \phi(d(x_i, x_h))) \quad (9)$$

As we said, the domain of the target variable is also a piece of evidence, so the domain mass (m^*) is obtained using (1) and resulting in,

$$m^* = 1 - \sum_{j=1}^N m_j(x_i) \quad (10)$$

Finally, to obtain the predicted value of the target variable y_i , we need to transform our belief function into a probability function, satisfying certain axiomatic requirements. Smets et al. showed that the Pignistic transformation could be used for this purpose. With this function, we can get the expected value of the predicted target variable y_i as:

$$\hat{y}_i = \sum_{j=1}^N m_j(x_i) \cdot y_j + \frac{m^* \cdot (\sup_{y \in \mathcal{L}} y + \inf_{y \in \mathcal{L}} y)}{2}$$

With lower and upper expectations:

$$\begin{aligned} \hat{y}_i^* &= \sum_{j=1}^N m_j(x_i) \cdot y_j + m^* \cdot \sup_{y \in \mathcal{L}} y \\ \hat{y}_{i*} &= \sum_{j=1}^N m_j(x_i) \cdot y_j + m^* \cdot \inf_{y \in \mathcal{L}} y \end{aligned}$$

4.2. Training phase: For a prediction to adjust as close as possible to its real value, we need to find the optimal hyperparameters of our

model. These parameters are in the discount function and the weights of each dimension used in the distance function. To obtain these values, we opted for using gradient-based algorithms for the training phase because they have a clear mathematical justification for reaching optimum values. The gradient descent algorithm is an iterative algorithm that optimizes variable values in order to minimize or maximize a target function. We used the Mean Absolute Error (MAE) as our target function, which is obtained:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n ||y_i - \hat{y}_i||$$

Where n is the number of patients, y_i is the true value of the future cost for patient i and \hat{y}_i is the predicted value i . Given the loss function MAE that depends on the variable v_t , which can be Γ or w , our goal is to minimize the value of the MAE function. The updated value of v_t called v_{t+1} is given by the following formula

$$v_{t+1} = v_t - \alpha \frac{\partial \text{MAE}}{\partial v}$$

Where α is called the learning rate, this algorithm gives us a sequence of values for v_0, \dots, v_k that minimizes the MAE, the initial value for v (i.e., v_0) is usually selected randomly. To apply gradient descent during the training phase, we try to predict the cost of each patient p_i in the training set L Using all the others $N - 1$ patients in L as evidence. An iteration computing predictions for all N patients is called an epoch, the method of the gradient descent converges by performing multiple epochs.

4.3. Computing Optimization:

The computation time of a prediction grows linearly with the size of the training set. To compute the mass of a vector of dimension m in a dataset with a training set of size n , first we need the discounting function; this has a complexity of $O(m)$. We can compute every discounting function of the input vector with the training set in $O(mn)$. Then we can obtain K , each product sequence $O(mn)$ and the summation also takes $O(mn)$, so we can compute K in time $O(mn)$. Finally to obtain the mass we need the discounting function, K , and a product sequence, so we compute the masses of the input vector, maintaining $O(mn)$ So then we can get the prediction using, with $O(mn)$ complexity.

Ref. has shown that it is possible to use a K -nearest Neighbors approach to speed up computation without a significant drop in performance. In particular, we used the implementation by Johnson et al. for the exact nearest

neighbour's search based on product Quantization. With this optimization, we create indexes for the K-nearest search during the training phase in time $O(mn + Kn)$. Then a new prediction only needs the masses of the K-Nearest Neighbours, which will be computed in $O(Km)$; the other masses are assumed to be null. Thus, the complexity for a prediction is $O(Km)$, once the model is trained.

4.4. INTERPRETABILITY:

The IEVREG is a transparent model. We can obtain the contribution (mass) of each piece of evidence in the training set L for every prediction we make. Thus, we fully know how the predicted value is computed. This model can already be considered as an interpretable one, but for the model to be fully explanatory, we will create a set of rules for each prediction with the masses obtained from the training set and the weights of each dimension learned during the training phase. The goal is to estimate the amount each set of evidence contributes to the prediction. First, we create a set of rules for each one of the patients in the training set using the masses of the other $N - 1$ patients in the sets the weights of the dimension. These rules encode the ranges of the dimensions for each of the input features and their masses. Then, for making a prediction, the model finds the most similar patients in the training set and combine their rules to create a new set of rules for that prediction. To illustrate how we obtain the rules with the IEVREG model we use a small health insurance dataset with only five dimensions as input. The five input data (dimensions) and the predicted value for the care costs are shown in the given table.

TABLE: 3

Age	Gender	BMI	Children	Smoker	True Score	Predicted Score
41	Male	32.2	2	No	6836	7555

We predicted the score of this patient using only the 50 nearest neighbours. Then, we obtain the most important rules (higher weights) for the cost prediction; these are shown in table4. These rules are the ranges and values that the patient shares with the training set patients:

Table 4. Rules.

Rule	Weight
$31.7640 < \text{BMI} < 32.9670$	0.48
gender = 0.0	0.48
smoker = 0.0	0.48
children = 2.0	0.48
$40.3107 < \text{age} < 41.3560$	0.22

In Table4 we can observe the interpretation of a patient's cost prediction (the one on Table3). The IEVREG model assigns low weight to age, and high weight to the other ones. As a consequence, the algorithm tries to find similar patients in terms of BMI, gender, smoker status, and children, and not worry much about age.

5.DICUSSION:

We provided a novel linear regression technique that can simply demonstrate the purposes for producing a certain forecast regarding potential healthcare expenses, which is a useful capacity in the medical field. We evaluated its journal of health engineering.

Table3: Details of patient's

Gender	BMI	Smoker	Age	Children	Actual value	Forecasted value
Female	29.98	No	37	1	6245	7154
Male	32.12	No	40	2	6725	7540

Gender	Estimated values	Weights
Male	30.6530 < BMI < 31.8560	0.45
	Gender 0.0	0.45
	Children 0.0	0.45
	Smoker 0.0	0.45
Female	39.2016 < age < 40.2451	0.22
	28.5421 < BMI < 29.7451	0.39
	Gender 0.0	0.39
	Children 0.0	0.39
	Smoker 0.0	0.39
	36.2016 < age < 37.2452	0.19

Require: Training data D , number of epochs e , learning rate η , and standard deviation σ .

Ensure: Weights. $\omega_0, \omega_1, \dots, \omega_k$

- (1) Initialize weights $\omega_0, \omega_1, \dots, \omega_k$ from standard normal distribution with zero mean and standard deviation σ .
 - for** epoch in $1, \dots, e$ **do**
 - for** each $(x, y) \in D$ in random order **do**
 - $\hat{y} \leftarrow \omega_0 + \sum_{i=1}^k \omega_i x_i$
 - if** $(\hat{y} > 1 \text{ and } y = 1)$ or $(\hat{y} < -1 \text{ and } y = -1)$ **then**
 - Continue**
 - $\omega_0 \leftarrow \omega_0 - \eta 2 (\hat{y} - y)$
 - for** i in $1, \dots, k$ **do**
 - $\omega_i \leftarrow \omega_i - \eta 2 (\hat{y} - y) x_i$
 - end for**
 - end for**
 - return** $\omega_0, \omega_1, \dots, \omega_k$

ALGORITHM 1: Linear regression (LR).

We evaluated its outcomes to the forecasting produced by the finest algorithms from the analysed research and reported to see how well it predicted. The linear regression is what we are talking about here. When we compare the outcomes of previous

designs for the cost of healthcare forecasting approach, we can see that our system is more efficient, demonstrating that a more explicit approach for an issue such as healthcare cost forecasting is conceivable. Our research, on the other hand, clearly reveals that healthcare spending is highly connected inside the Medicare program. There are approximately numerous people enrolled in the program. This finding could lead to preventive measures. Autocorrelation shows an inherent methodology that could be influenced by variables that can be changed. As a result, clinicians can use more accurate machine-learning algorithms to target these therapies to the proper HCHN group. There are a few flaws in this research. Initially, we performed the research within the context of a single state's Medicare system. The outcomes might differ depending on the state or kind of payer. Secondly, only general-purpose machine-learning algorithms were used. Certain customized versions might function optimally. Thirdly, the prediction algorithms offer no direction on the preventive characteristics that should be considered when developing treatments. Lastly, determining overall health solely based on claim statistics is restricted. Further input resources, such as descriptive elements of electronic health records (EHRs), illness intensity assessments, and socioeconomic determinants of health care, might well be required. A few of these restrictions will be addressed in the forthcoming research

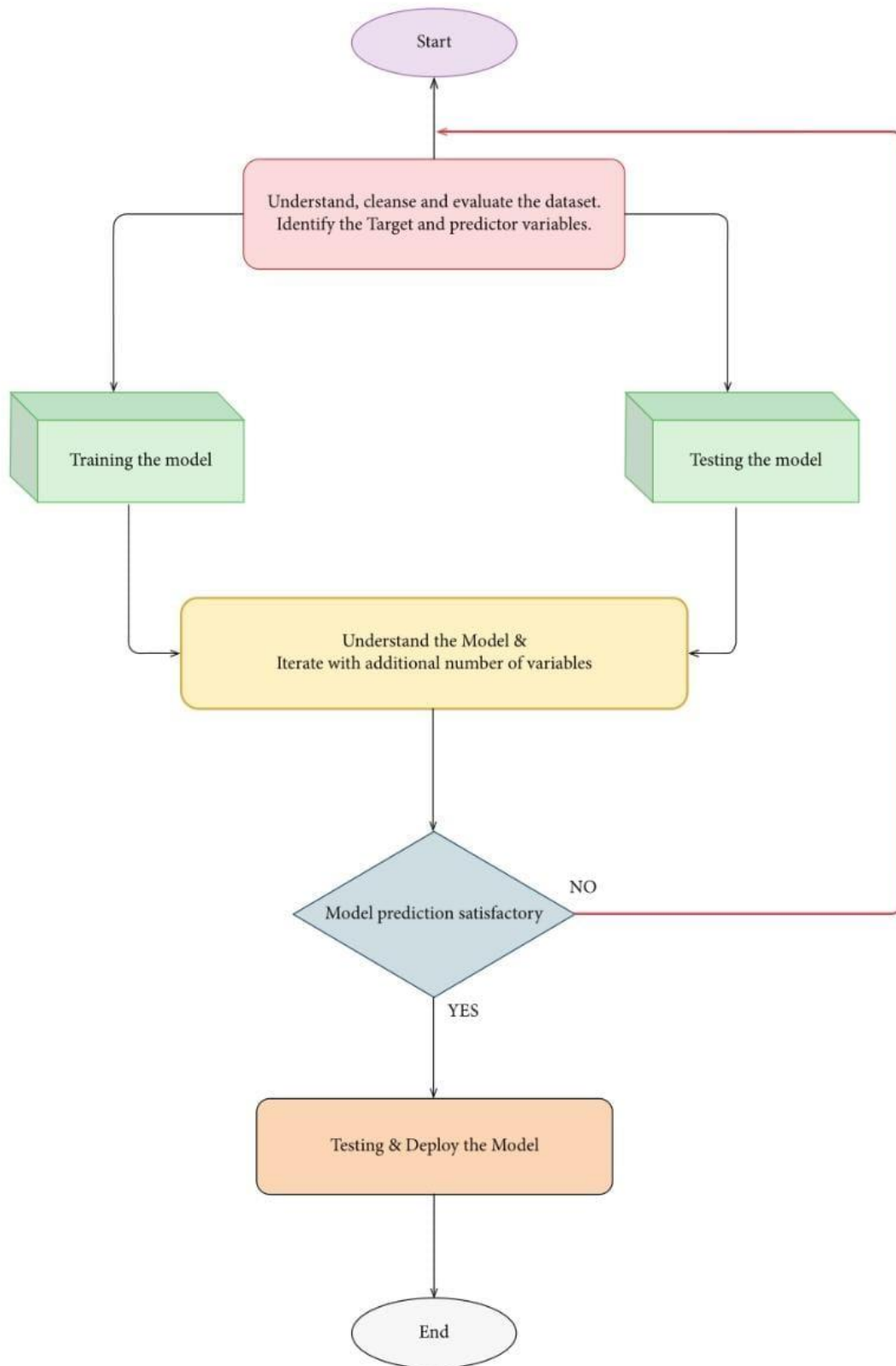


FIGURE 3: Flowchart for estimating the healthcare costs.

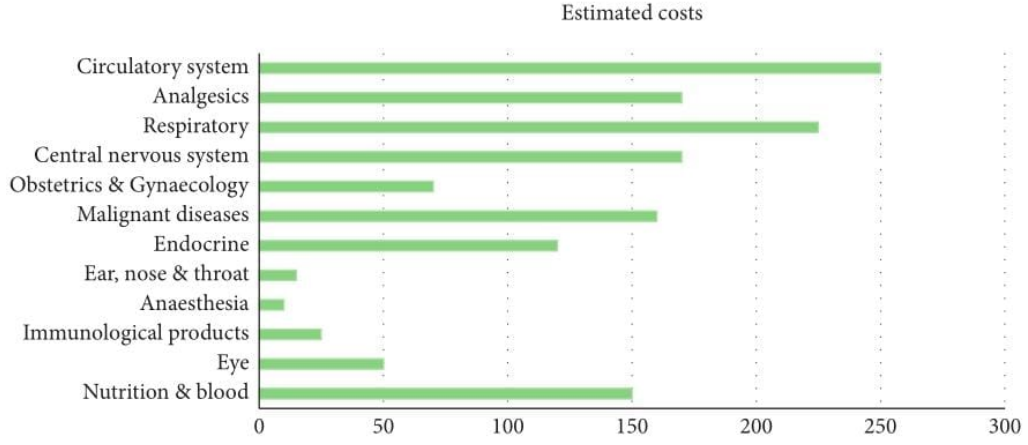


FIGURE 4: Healthcare expenses attributable to obesity and overweight between people on a yearly basis.

- EXPERIMENTS AND RESULTS:

To evaluate the performance of our model, we compare its results with two other methods reported by Morid et al. and Duncan et al. for the health cost prediction problem; these works used GB and ANN methods respectively. To measure the performance of each method, we may use the MAE, which computes the average absolute difference between the predicted cost and the real one. However, the MAE is not useful to compare results with costs expressed in different currencies, so we will also use the Mean Absolute Percentage Error (MAPE) a modified absolute error where the MAE is divided by the mean cost and is computed as

$$MAPE = \frac{\frac{1}{n} \sum_{i=1}^n ||y_i - \hat{y}_i||}{\bar{m}}$$

6.DICUSSION:

We provided a new linear regression that can easily demonstrate the reasons for producing a certain forecast regarding potential healthcare expenses, which is a useful capacity in the healthcare area. The linear regression algorithm is used to estimate the healthcare costs of the patients such as obesity (BMI) using certain devices such as smartphones and smart devices. For estimation, by the use of linear regression, supervised learning performs more accurately. By providing comprehensive evidence, regression methodology can be effectively used for prognosis in conjunction with the dataset. The domain and time accuracy will determine the prediction model and the estimation of healthcare expenses. The proposed method reduces the risk of overfitting, and also, training time is less. This method is effective in estimating the healthcare costs of patients with an accuracy rate of 97.89%. The extensive tests on a real-time world database have confirmed the efficiency of our method.

Data Availability:

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest:

The authors declare that there are no conflicts of interest regarding the publication of this study.

DISCUSSION:

We presented a new regression method that has the ability to easily show the reasons for making a particular prediction about possible health care costs, which is a desirable ability in the health domain. In order to test its predicting performance, we compared its results with the predictions made by two best models from the eleven analyzed and reported by Morid et al.. This is the GB and the ANN. Comparing the results for the three models for the health care cost prediction problem we can conclude that our method obtains better performance, proving that it is possible to create a more transparent model for a problem like health care cost prediction, going against the common belief that sophisticated and black-box like methods are always the solution with the best performance for every problem that is presented. We improved the Evidence regression model presented by Petit-Renaud and Denoeux to be used in the prediction of health care costs. Our results obtained using data of electronic health records from Tsuyama Chuo Hospital showed that our

Evidence regression model with an $R^2 = 0.44$, a transparent and interpretable method, could outperformed the current state of the art supervised learning algorithms such as GB ($R^2 = 0.40$) and ANN ($R^2 = 0.35$). Even though results are similar or better than other previous works, we believe our results are still improvable. One of the approaches to improve performance was classifying patients in cost buckets as recommended by various studies, this resulted in better performances but escaped the goal of this work, so for future work we can apply this classification process to obtain a patient risk class as first step to improve the performance of our IEVREG model, and to continue comparing our model to even more sophisticated methods we could try to solve the prediction of health care cost using deep learning methods but for this to be feasible we need a larger dataset. We also plan to apply this model to other regression problems in the health care domain, for example, predicting the hospital length of stay and predicting the days of readmission based on each patient's diagnosis and history, which are two classic prediction problems for this domain.

CODE:

```
<!DOCTYPE html>
<html lang="en">

<head>
  <meta charset="utf-8">
  <meta content="width=device-width, initial-scale=1.0" name="viewport">

  <title>Medistats: Unveiling Insights, Transforming Healthcare Costs</title>
  <meta content="" name="description">
  <meta content="" name="keywords">

  <link href="static/assets/img/favicon.png" rel="icon">
  <link href="static/assets/img/apple-touch-icon.png" rel="apple-touch-
icon">

  <link
href="https://fonts.googleapis.com/css?family=Open+Sans:300,300i,400,400
i,600,600i,700,700i|Roboto:300,300i,400,400i,500,500i,600,600i,700,700i|Po
ppins:300,300i,400,400i,500,500i,600,600i,700,700i" rel="stylesheet">

  <link href="static/assets/vendor/fontawesome-free/css/all.min.css"
rel="stylesheet">
  <link href="static/assets/vendor/animate.css/animate.min.css"
rel="stylesheet">
  <link href="static/assets/vendor/aos/aos.css" rel="stylesheet">
  <link href="static/assets/vendor/bootstrap/css/bootstrap.min.css"
rel="stylesheet">
  <link href="static/assets/vendor/bootstrap-icons/bootstrap-icons.css"
rel="stylesheet">
  <link href="static/assets/vendor/boxicons/css/boxicons.min.css"
rel="stylesheet">
  <link href="static/assets/vendor/glightbox/css/glightbox.min.css"
rel="stylesheet">
  <link href="static/assets/vendor/swiper/swiper-bundle.min.css"
rel="stylesheet">

  <link href="static/assets/css/style.css" rel="stylesheet">

</head>
```

```

<body>

<div id="topbar" class="d-flex align-items-center fixed-top">
  <div class="container d-flex align-items-center justify-content-center
justify-content-md-between">
    <div class="align-items-center d-none d-md-flex">
      <i class="bi bi-clock"></i> Unveiling Insights, Transforming Healthcare
Costs
    </div>
    <div class="d-flex align-items-center">
      <i class="bi bi-phone"></i> Version 1.0
    </div>
  </div>
</div>

<header id="header" class="fixed-top">
  <div class="container d-flex align-items-center">

    <a href="index.html" class="logo me-auto"></a>

    <!-- <h1 class="logo me-auto"><a href="index.html">Medistats</a></h1>
-->

    <nav id="navbar" class="navbar order-last order-lg-0">
      <ul>
        <li><a class="nav-link scrollto " href="#hero">Home</a></li>
        <li><a class="nav-link scrollto" href="#about">About</a></li>
        <li><a class="nav-link scrollto" href="#features">Features</a></li>
        <li><a class="nav-link scrollto"
href="#appointment">Dashboard</a></li>
        <li><a class="nav-link scrollto" href="#services">Story</a></li>
        <li><a class="nav-link scrollto" href="#analysis">OR Analysis</a></li>
        <li><a class="nav-link scrollto" href="#gallery">Gallery</a></li>
        <li><a class="nav-link scrollto" href="#faq">FAQ</a></li>

      </ul>
      <i class="bi bi-list mobile-nav-toggle"></i>
    </nav><!-- .navbar -->

    <a href="#gallery" class="appointment-btn scrollto"><span class="d-none
d-md-inline">Visuals</span></a>

```



```

</div>
</header><!-- End Header -->

<section id="hero">
  <div id="heroCarousel" data-bs-interval="5000" class="carousel slide
carousel-fade" data-bs-ride="carousel">

    <ol class="carousel-indicators" id="hero-carousel-indicators"></ol>

    <div class="carousel-inner" role="listbox">

      <div class="carousel-item active" style="background-image:
url(static/assets/img/slide/slide-1.jpg)">
        <div class="container">
          <h2>Welcome to <span>Medistats</span></h2>
          <p>Welcome to our project on the <b>Estimation and Prediction of
Hospitalization and Medical Care Costs!</b> We leverage advanced data
analytics techniques to uncover the factors influencing medical expenses.
</p>
          <a href="#about" class="btn-get-started scrollto">Read More</a>
        </div>
      </div>

      <div class="carousel-item" style="background-image:
url(static/assets/img/slide/slide-2.jpg)">
        <div class="container">
          <h2>Compelling Visualizations</h2>
          <p>Through interactive <b>data visualizations</b>, we provide
insights into the relationships between variables such as age, gender, BMI,
smoking habits, and region, enabling informed decision-making. Join us on
this data-driven journey to unravel the complexities of healthcare costs and
contribute to the advancement of healthcare knowledge. Explore, analyze, and
predict medical care expenses with our powerful platform.</p>
          <a href="#about" class="btn-get-started scrollto">Read More</a>
        </div>
      </div>

      <div class="carousel-item" style="background-image:
url(static/assets/img/slide/slide-3.jpg)">
        <div class="container">

```

```
        <h2>Come and Explore</h2>
        <p>Join us on this <b>data-driven journey</b> to unravel the
        complexities of healthcare costs and contribute to the advancement of
        healthcare knowledge. Explore, analyze, and predict medical care expenses
        with our powerful platform.</p>
        <a href="#about" class="btn-get-started scrollto">Read More</a>
    </div>
</div>
```

```
    <a class="carousel-control-prev" href="#heroCarousel" role="button"
    data-bs-slide="prev">
        <span class="carousel-control-prev-icon bi bi-chevron-left" aria-
        hidden="true"></span>
    </a>
```

```
    <a class="carousel-control-next" href="#heroCarousel" role="button"
    data-bs-slide="next">
        <span class="carousel-control-next-icon bi bi-chevron-right" aria-
        hidden="true"></span>
    </a>
```

```
</div>
</section><!-- End Hero -->
```

```
<main id="main">
```

```
<!-- ===== Featured Services Section ===== -->
<section id="featured-services" class="featured-services">
    <div class="container" data-aos="fade-up">

        <div class="row">
            <div class="col-md-6 col-lg-3 d-flex align-items-stretch mb-5 mb-lg-0">
                <div class="icon-box" data-aos="fade-up" data-aos-delay="100">
                    <div class="icon"><i class="fas fa-heartbeat"></i></div>
                    <h4 class="title"><a href="">UNCOVER</a></h4>
                    <p class="description">the financial implications of various medical
                    conditions on and beyond the heart.</p>
                </div>
            </div>

            <div class="col-md-6 col-lg-3 d-flex align-items-stretch mb-5 mb-lg-0">
```

```

<div class="icon-box" data-aos="fade-up" data-aos-delay="200">
  <div class="icon"><i class="fas fa-pills"></i></div>
  <h4 class="title"><a href="">GAIN</a></h4>
  <p class="description">valuable financial insights into the medical
expense realm.</p>
</div>
</div>

```

```

<div class="col-md-6 col-lg-3 d-flex align-items-stretch mb-5 mb-lg-0">
  <div class="icon-box" data-aos="fade-up" data-aos-delay="300">
    <div class="icon"><i class="fas fa-thermometer"></i></div>
    <h4 class="title"><a href="">EMPOWER</a></h4>
    <p class="description">individuals with proactive financial planning
by providing accurate cost projections.</p>
  </div>
</div>

```

```

<div class="col-md-6 col-lg-3 d-flex align-items-stretch mb-5 mb-lg-0">
  <div class="icon-box" data-aos="fade-up" data-aos-delay="400">
    <div class="icon"><i class="fas fa-dna"></i></div>
    <h4 class="title"><a href="">SAVE</a></h4>
    <p class="description">your time by making quick-informed decisions
through inferencing our visualizations.</p>
  </div>
</div>

```

```

</div>

```

```

</div>

```

```

</section><!-- End Featured Services Section -->

```

```

<!-- ===== Cta Section ===== -->

```

```

<section id="cta" class="cta">

```

```

  <div class="container" data-aos="zoom-in">

```

```

    <div class="text-center">

```

```

      <h3>In a Hurry? Dive into our Visuals!</h3>

```

```

      <p> Discover the power of data visualization instantly. Our project offers
intuitive and interactive visuals for quick insights into medical costs. Explore
charts, graphs, and maps to make informed decisions in no time. Experience
the convenience of understanding complex data at a glance.</p>

```

```

      <a class="cta-btn scrollto" href="#appointment">Let's Go</a>

```

```

    </div>

```

```
</div>
</section><!-- End Cta Section -->
```

```
<!-- ===== About Us Section ===== -->
<section id="about" class="about">
  <div class="container" data-aos="fade-up">
```

```
    <div class="section-title">
      <h2>About Us</h2>
      <p>Welcome to our project on the Estimation and Prediction of
Hospitalization and Medical Care Costs! We are a team of dedicated
individuals passionate about exploring the intersection of healthcare and data
analysis.</p>
    </div>
```

```
    <div class="row">
      <div class="col-lg-6" data-aos="fade-right">
        <br><br><br>
        
```

```
      </div>
      <div class="col-lg-6 pt-4 pt-lg-0 content" data-aos="fade-left">
        <h3>Team Introduction:</h3>
        <p class="fst-italic">
          Allow us to introduce ourselves:
        </p>
        <ul>
          <li><i class="bi bi-check-circle"></i><b>Aditya:</b> Hi, I'm Aditya!
With an interest in statistics, I bring my passion in analyzing complex datasets
and extracting meaningful insights. I'm excited to dive deep into the realm of
healthcare data to uncover patterns and trends related to medical costs.</li>
          <li><i class="bi bi-check-circle"></i><b>Divyansh:</b> Hello, I'm
Divyansh! As a data scientist, I thrive on transforming raw data into
actionable information. I'm fascinated by the possibilities of using data
analytics to understand the factors influencing hospitalization and medical
care costs. Let's discover the hidden stories within the numbers!</li>
          <li><i class="bi bi-check-circle"></i><b>Sutanuka:</b> Greetings, I'm
Sutanuka! My passion lies in healthcare research and policy analysis. By
combining my knowledge of healthcare systems with data-driven approaches,
I aim to contribute to a better understanding of cost estimation and prediction
in medical care. Let's make a positive impact on healthcare outcomes!</li>
```

Rishi: Hi there, I'm Rishi! I specialize in data visualization and user experience design. I believe in the power of visual storytelling to communicate complex information effectively. Through intuitive and insightful visualizations, I strive to bring clarity and engagement to our project.

Together, we form a dynamic team, united by our shared goal of unraveling the complexities of hospitalization and medical care costs. Our project aims to provide valuable insights into the factors influencing medical expenses, enabling individuals and organizations to make informed decisions and improve healthcare outcomes.

Records were analysed

<https://www.kaggle.com/datasets/mirichoi0218/insurance> Find out more »

```
<p><strong>Members</strong> organized the project</p>
<a href="#about">Find out more &raquo;</a> <!-- About Us Section-->
</div>
</div>
```

```
<div class="col-lg-3 col-md-6 d-md-flex align-items-md-stretch">
  <div class="count-box">
    <i class="fas fa-flask"></i>
    <span data-purecounter-start="0" data-purecounter-end="14" data-
purecounter-duration="1" class="purecounter"></span>
    <p><strong>interactive</strong> visuals</p>
    <a href="#gallery">Find out more &raquo;</a>
  </div>
</div>
```

```
<div class="col-lg-3 col-md-6 d-md-flex align-items-md-stretch">
  <div class="count-box">
    <i class="fas fa-award"></i>
    <span data-purecounter-start="0" data-purecounter-end="4" data-
purecounter-duration="1" class="purecounter"></span>
    <p><strong>technologies</strong> implemented</p>
    <a href="#"
onclick="window.open('https://flask.palletsprojects.com/en/2.3.x/');
window.open('https://help.tableau.com/current/pro/desktop/en-
us/gettingstarted_overview.htm');window.open('https://dev.mysql.com/doc
/');window.open('https://getbootstrap.com/docs/4.1/getting-
started/introduction/');">Find out more &raquo;</a> <!-- Documentation
links. Won't work with pop-up blockers-->
  </div>
</div>
```

```
</div>

</div>
</section><!-- End Counts Section -->
```

```
<section id="features" class="features">
  <div class="container" data-aos="fade-up">
```

```
    <div class="row">
      <div class="col-lg-6 order-2 order-lg-1" data-aos="fade-right">
        <div class="icon-box mt-5 mt-lg-0">
          <i class="bx bx-receipt"></i>
```

```
<h4>Interactive Visualizations</h4>
<p>Explore dynamic charts, graphs, and maps that allow you to
interact with medical cost data, enabling a deeper understanding of patterns
and relationships.</p>
</div>
<div class="icon-box mt-5">
  <i class="bx bx-cube-alt"></i>
  <h4>Customizable Views</h4>
  <p>Tailored visualizations to your needs by adjusted parameters,
filters, and variables, focusing on specific aspects of medical costs for
personalized insights.</p>
</div>
<div class="icon-box mt-5">
  <i class="bx bx-images"></i>
  <h4>Comparative Analysis</h4>
  <p>Compare scenarios, and variables using visualizations that
facilitate side-by-side analysis, helping identify trends, patterns, and
outliers.</p>
</div>
<div class="icon-box mt-5">
  <i class="bx bx-shield"></i>
  <h4>Storytelling Capabilities</h4>
  <p>Experience engaging visual narratives that combine visual
elements, annotations, and narratives to communicate complex medical cost
information effectively.</p>
</div>
</div>
<div class="image col-lg-6 order-1 order-lg-2" style='background-image:
url("static/assets/img/features.jpg");' data-aos="zoom-in"></div>
</div>
</section><!-- End Features Section -->
```

quisquam cupiditate. Et nemo qui impedit suscipit alias ea. Quia fugiat sit in iste officiis commodi quidem hic quas.</p>
</div>

```
<div class="row" data-aos="fade-up" data-aos-delay="100">
  <div class="col-lg-4 mb-5 mb-lg-0">
    <ul class="nav nav-tabs flex-column">
      <li class="nav-item">
        <a class="nav-link active show" data-bs-toggle="tab" data-bs-
target="#tab-1">
          <h4>Cardiology</h4>
          <p>Quis excepturi porro totam sint earum quo nulla perspiciatis
eius.</p>
        </a>
      </li>
      <li class="nav-item mt-2">
        <a class="nav-link" data-bs-toggle="tab" data-bs-target="#tab-2">
          <h4>Neurology</h4>
          <p>Voluptas vel esse repudiandae quo excepturi.</p>
        </a>
      </li>
      <li class="nav-item mt-2">
        <a class="nav-link" data-bs-toggle="tab" data-bs-target="#tab-3">
          <h4>Hepatology</h4>
          <p>Velit veniam ipsa sit nihil blanditiis mollitia natus.</p>
        </a>
      </li>
      <li class="nav-item mt-2">
        <a class="nav-link" data-bs-toggle="tab" data-bs-target="#tab-4">
          <h4>Pediatrics</h4>
          <p>Ratione hic sapiente nostrum doloremque illum nulla
praesentium id</p>
        </a>
      </li>
    </ul>
  </div>
  <div class="col-lg-8">
    <div class="tab-content">
      <div class="tab-pane active show" id="tab-1">
        <h3>Cardiology</h3>
        <p class="fst-italic">Qui laudantium consequatur laborum sit qui ad
sapiente dila parde sonata raqer a videna mareta paulona marka</p>
```


<p>Et nobis maiores eius. Voluptatibus ut enim blanditiis atque harum sint. Laborum eos ipsum ipsa odit magni. Incidunt hic ut molestiae aut qui. Est repellat minima eveniet eius et quis magni nihil. Consequatur dolorem quaerat quos qui similique accusamus nostrum rem vero</p>

</div>

<div class="tab-pane" id="tab-2">

<h3>Neurology</h3>

<p class="fst-italic">Qui laudantium consequatur laborum sit qui ad sapiente dila parde sonata raquer a videna mareta paulona marka</p>

<p>Et nobis maiores eius. Voluptatibus ut enim blanditiis atque harum sint. Laborum eos ipsum ipsa odit magni. Incidunt hic ut molestiae aut qui. Est repellat minima eveniet eius et quis magni nihil. Consequatur dolorem quaerat quos qui similique accusamus nostrum rem vero</p>

</div>

<div class="tab-pane" id="tab-3">

<h3>Hepatology</h3>

<p class="fst-italic">Qui laudantium consequatur laborum sit qui ad sapiente dila parde sonata raquer a videna mareta paulona marka</p>

<p>Et nobis maiores eius. Voluptatibus ut enim blanditiis atque harum sint. Laborum eos ipsum ipsa odit magni. Incidunt hic ut molestiae aut qui. Est repellat minima eveniet eius et quis magni nihil. Consequatur dolorem quaerat quos qui similique accusamus nostrum rem vero</p>

</div>

<div class="tab-pane" id="tab-4">

<h3>Pediatrics</h3>

<p class="fst-italic">Qui laudantium consequatur laborum sit qui ad sapiente dila parde sonata raquer a videna mareta paulona marka</p>

<p>Et nobis maiores eius. Voluptatibus ut enim blanditiis atque harum sint. Laborum eos ipsum ipsa odit magni. Incidunt hic ut molestiae aut qui. Est repellat minima eveniet eius et quis magni nihil. Consequatur dolorem quaerat quos qui similique accusamus nostrum rem vero</p>

</div>

</div>

</div>

</div>

</div>

</section> End Departments Section -->

<!-- ===== Testimonials Section =====

<section id="testimonials" class="testimonials">

<div class="container" data-aos="fade-up">

<div class="section-title">

<h2>Testimonials</h2>

<p>Magnam dolores commodi suscipit. Necessitatibus eius consequatur ex aliquid fuga eum quidem. Sit sint consectetur velit. Quisquam quos quisquam cupiditate. Et nemo qui impedit suscipit alias ea. Quia fugiat sit in iste officiis commodi quidem hic quas.</p>

</div>

<div class="testimonials-slider swiper" data-aos="fade-up" data-aos-delay="100">

<div class="swiper-wrapper">

<div class="swiper-slide">

<div class="testimonial-item">

<p>

<i class="bx bxs-quote-alt-left quote-icon-left"></i>

Proin iaculis purus consequat sem cure digni ssim donec porttitora entum suscipit rhoncus. Accusantium quam, ultricies eget id, aliquam eget nibh et. Maecen aliquam, risus at semper.

<i class="bx bxs-quote-alt-right quote-icon-right"></i>

</p>

<h3>Saul Goodman</h3>

<h4>CEO & Founder</h4>

</div>

</div> End testimonial item

<div class="swiper-slide">

<div class="testimonial-item">

<p>

<i class="bx bxs-quote-alt-left quote-icon-left"></i>

Export tempor illum tamen malis malis eram quae irure esse labore quem cillum quid cillum eram malis quorum velit fore eram velit sunt aliqua noster fugiat irure amet legam anim culpa.

```

        <i class="bx bxs-quote-alt-right quote-icon-right"></i>
    </p>
    
    <h3>Sara Wilsson</h3>
    <h4>Designer</h4>
</div>
</div> End testimonial item

<div class="swiper-slide">
    <div class="testimonial-item">
        <p>
            <i class="bx bxs-quote-alt-left quote-icon-left"></i>
            Enim nisi quem export dui labore cillum quae magna enim sint
            quorum nulla quem veniam dui minim tempor labore quem eram dui noster
            aute amet eram fore quis sint minim.
            <i class="bx bxs-quote-alt-right quote-icon-right"></i>
        </p>
        
        <h3>Jena Karlis</h3>
        <h4>Store Owner</h4>
    </div>
</div> End testimonial item

<div class="swiper-slide">
    <div class="testimonial-item">
        <p>
            <i class="bx bxs-quote-alt-left quote-icon-left"></i>
            Fugiat enim eram quae cillum dolore dolor amet nulla culpa multos
            export minim fugiat minim velit minim dolor enim dui veniam ipsum anim
            magna sunt elit fore quem dolore labore illum veniam.
            <i class="bx bxs-quote-alt-right quote-icon-right"></i>
        </p>
        
        <h3>Matt Brandon</h3>
        <h4>Freelancer</h4>
    </div>
</div> End testimonial item

<div class="swiper-slide">
    <div class="testimonial-item">

```

```
<p>
  <i class="bx bxs-quote-alt-left quote-icon-left"></i>
  Quis quorum aliqua sint quem legam fore sunt eram irure aliqua
veniam tempor noster veniam enim culpa labore dui sunt culpa nulla illum
cillum fugiat legam esse veniam culpa fore nisi cillum quid.
```

```
  <i class="bx bxs-quote-alt-right quote-icon-right"></i>
</p>

  <h3>John Larson</h3>
  <h4>Entrepreneur</h4>
</div>
</div> End testimonial item
```

```
</div>
<div class="swiper-pagination"></div>
</div>
```

```
</div>
</section> End Testimonials Section -->
```

```
<!-- ===== Doctors Section =====
<section id="doctors" class="doctors section-bg">
  <div class="container" data-aos="fade-up">

    <div class="section-title">
      <h2>Doctors</h2>
      <p>Magnam dolores commodi suscipit. Necessitatibus eius consequatur
ex aliquid fuga eum quidem. Sit sint consectetur velit. Quisquam quos
quisquam cupiditate. Et nemo qui impedit suscipit alias ea. Quia fugiat sit in
iste officiis commodi quidem hic quas.</p>
    </div>
```

```
  <div class="row">

    <div class="col-lg-3 col-md-6 d-flex align-items-stretch">
      <div class="member" data-aos="fade-up" data-aos-delay="100">
        <div class="member-img">
          
          <div class="social">
            <a href=""><i class="bi bi-twitter"></i></a>
            <a href=""><i class="bi bi-facebook"></i></a>
```

```

        <a href=""><i class="bi bi-instagram"></i></a>
        <a href=""><i class="bi bi-linkedin"></i></a>
    </div>
</div>
<div class="member-info">
    <h4>Walter White</h4>
    <span>Chief Medical Officer</span>
</div>
</div>
</div>

<div class="col-lg-3 col-md-6 d-flex align-items-stretch">
    <div class="member" data-aos="fade-up" data-aos-delay="200">
        <div class="member-img">
            
            <div class="social">
                <a href=""><i class="bi bi-twitter"></i></a>
                <a href=""><i class="bi bi-facebook"></i></a>
                <a href=""><i class="bi bi-instagram"></i></a>
                <a href=""><i class="bi bi-linkedin"></i></a>
            </div>
        </div>
        <div class="member-info">
            <h4>Sarah Jhonson</h4>
            <span>Anesthesiologist</span>
        </div>
    </div>
</div>

<div class="col-lg-3 col-md-6 d-flex align-items-stretch">
    <div class="member" data-aos="fade-up" data-aos-delay="300">
        <div class="member-img">
            
            <div class="social">
                <a href=""><i class="bi bi-twitter"></i></a>
                <a href=""><i class="bi bi-facebook"></i></a>
                <a href=""><i class="bi bi-instagram"></i></a>
                <a href=""><i class="bi bi-linkedin"></i></a>
            </div>
        </div>
        <div class="member-info">

```

```
<h4>William Anderson</h4>
<span>Cardiology</span>
</div>
</div>
</div>
```

```
<div class="col-lg-3 col-md-6 d-flex align-items-stretch">
  <div class="member" data-aos="fade-up" data-aos-delay="400">
    <div class="member-img">
      
      <div class="social">
        <a href=""><i class="bi bi-twitter"></i></a>
        <a href=""><i class="bi bi-facebook"></i></a>
        <a href=""><i class="bi bi-instagram"></i></a>
        <a href=""><i class="bi bi-linkedin"></i></a>
      </div>
    </div>
    <div class="member-info">
      <h4>Amanda Jepson</h4>
      <span>Neurosurgeon</span>
    </div>
  </div>
</div>
```

```
</div>
```

```
</div>
```

```
</section> End Doctors Section -->
```

```
<!-- ===== Gallery Section ===== -->
```

```
<section id="gallery" class="gallery">
  <div class="container" data-aos="fade-up">
```

```
    <div class="section-title">
```

```
      <h2>Gallery</h2>
```

```
      <p>A collection of interactive visuals showcasing the estimation and
prediction of hospitalization and medical care costs. Explore insightful charts,
graphs, and maps that provide a deeper understanding of healthcare
expenses.</p>
```

```
    </div>
```

```
    <div class="gallery-slider swiper">
```

```
<div class="swiper-wrapper align-items-center">
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-1.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-2.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-3.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-4.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-5.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-6.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-7.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-8.jpg"></a></div>
  <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery/gallery-9.jpg"></a></div>
</div>
<div class="swiper-pagination"></div>
</div>

</div>
</section><!-- End Gallery Section -->
```

```

<!-- ===== Dashboard Section ===== -->
<section id="appointment" class="appointment section-bg">
  <div class="container" data-aos="fade-up">

    <div class="section-title">
      <h2>Dashboard</h2>
      <p>Explore our comprehensive dashboards for insightful analysis of hospitalization and medical care costs. Customize and filter data, uncover trends, and make data-driven decisions. Experience the power of visualizations in driving informed strategies and improving healthcare outcomes.</p>
      <div class='tableauPlaceholder' id='viz1688039609529' style='position: relative'>
        <noscript>
          <a href='#'>
            <img alt='Dashboard 4 '
src='https://public.tableau.com/static/images/DA&
#47;DAproject_16879674240780&#47;Dashboard4&#47;1_rss.png'
style='border: none' />
          </a>
        </noscript>
        <object class='tableauViz' style='display:none;'>
          <param name='host_url'
value='https%3A%2F%2Fpublic.tableau.com%2F' />
          <param name='embed_code_version' value='3' />
          <param name='site_root' value='' />
          <param name='name'
value='DAproject_16879674240780&#47;Dashboard4' />
          <param name='tabs' value='no' />
          <param name='toolbar' value='yes' />
          <param name='static_image'
value='https://public.tableau.com/static/images/D
A&#47;DAproject_16879674240780&#47;Dashboard4&#47;1.png' />
          <param name='animate_transition' value='yes' />
          <param name='display_static_image' value='yes' />
          <param name='display_spinner' value='yes' />
          <param name='display_overlay' value='yes' />
          <param name='display_count' value='yes' />
          <param name='language' value='en-US' />
        </object>
      </div>
    </div>
  </div>

```



```

<script type='text/javascript'>
  var divElement = document.getElementById('viz1688039609529');
  var vizElement = divElement.getElementsByTagName('object')[0];
  if (divElement.offsetWidth > 800) {
    vizElement.style.width = '1000px';
    vizElement.style.height = '827px';
  } else if (divElement.offsetWidth > 500) {
    vizElement.style.width = '1000px';
    vizElement.style.height = '827px';
  } else {
    vizElement.style.width = '100%';
    vizElement.style.height = '877px';
  }
  var scriptElement = document.createElement('script');
  scriptElement.src =
'https://public.tableau.com/javascripts/api/viz_v1.js';
  vizElement.parentNode.insertBefore(scriptElement, vizElement);
</script>
</div>

</div>
</section><!-- End Dashboard Section -->
<!-- ===== Story Section ===== -->
<section id="services" class="services services">
  <div class="section-title">
    <h2>Story</h2>
    <p>Embark on a captivating journey through our Stories section, where
we bring medical cost analysis to life. Dive into compelling narratives that
blend data visualizations and storytelling techniques, offering unique
perspectives and deep insights into the complex world of healthcare
expenses.</p>
    <div class='tableauPlaceholder' id='viz1688039896287' style='position:
relative'>
      <noscript>
        <a href='#'>
          <img alt='Story '
src='https://public.tableau.com/static/images/YB&
#47;YBBWGHRY8&47;1_rss.png' style='border: none' />
        </a>
      </noscript>
      <object class='tableauViz' style='display:none;'>
        <param name='host_url'
value='https%3A%2F%2Fpublic.tableau.com%2F' />

```

```

    <param name='embed_code_version' value='3' />
    <param name='path' value='shared&#47;YBBWGHRY8' />
    <param name='toolbar' value='yes' />
    <param name='static_image'
value='https:&#47;&#47;public.tableau.com&#47;static&#47;images&#47;Y
B&#47;YBBWGHRY8&#47;1.png' />
    <param name='animate_transition' value='yes' />
    <param name='display_static_image' value='yes' />
    <param name='display_spinner' value='yes' />
    <param name='display_overlay' value='yes' />
    <param name='display_count' value='yes' />
    <param name='language' value='en-US' />
  </object>
</div>
<script type='text/javascript'>
  var divElement = document.getElementById('viz1688039896287');
  var vizElement = divElement.getElementsByTagName('object')[0];
  vizElement.style.width = '100%';
  vizElement.style.height = (divElement.offsetWidth * 0.75) + 'px';
  var scriptElement = document.createElement('script');
  scriptElement.src =
'https://public.tableau.com/javascripts/api/viz_v1.js';
  vizElement.parentNode.insertBefore(scriptElement, vizElement);
</script>
</div>
</section>

```

```

<!-- Analysis section -->
<section id="analysis" class="services services">
  <div class="section-title">
    <h2>Odds Ratio Analysis</h2>
    <p>Introducing Odds Ratio Analysis: Unlock the Secrets of Healthcare
Costs. Delve into our cutting-edge feature that explores the intricate
relationships between BMI, smoking, sex, and medical charges. Uncover
hidden patterns and gain valuable insights to make informed decisions.
Harness the power of odds and revolutionize your understanding of
healthcare economics.</p>
    <br>
  </div>
  <div class="gallery-slider swiper">
    <div class="swiper-wrapper align-items-center">
      <div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery 1/gallery-1.jpg"></a></div>
```

```
<div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery 1/gallery-2.jpg"></a></div>
```

```
<div class="swiper-slide"><a class="gallery-lightbox"
href="static/assets/img/gallery 1/gallery-3.jpg"></a></div>
```

```
</div>
```

```
<br><br><br>
```

```
<div class="swiper-pagination"></div>
```

```
</div>
```

```
<div class="section-title">
```

```
<h4>Our cutting-edge Odds Ratio Analysis reveals remarkable
findings!</h4>
```

```
<div class="container" data-aos="fade-up">
```

```
<div class="row">
```

```
<div class="col-lg-6 order-2 order-lg-1" data-aos="fade-right">
```

```
<div class="icon-box mt-5 mt-lg-0">
```

```
<i class="bx bx-receipt"></i>
```

```
<h4>Odds Ratio for BMI</h4>
```

```
<p>BMI showcases a compelling relationship, with an Odds Ratio of
0.0485 (95% CI: 0.0359 - 0.0656).</p>
```

```
</div>
```

```
<div class="icon-box mt-5">
```

```
<i class="bx bx-cube-alt"></i>
```

```
<h4>Odds Ratio for Sex</h4>
```

```
<p>Sex, while not exhibiting significant influence (Odds Ratio:
1.1301, 95% CI: 0.8969 - 1.4239), provides valuable contextual insights.</p>
```

```
</div>
```

```
<div class="icon-box mt-5">
```

```
<i class="bx bx-images"></i>
```

```
<h4>Odds Ratio for Smoking</h4>
```

```
<p>Smoking emerges as the dominant factor, with an astounding
Odds Ratio of 1703 (95% CI: 237.1882 - 12227.4568).</p>
```

```
</div>
```

```
</div>
```

```
<div class="image col-lg-6 order-1 order-lg-2" style='background-
image: url("static/assets/img/features 1.jpg");' data-aos="zoom-in"></div>
```

```

    </div>
    <br>
</div>
<br><br>
<div class="section-title">
    <h4>Say No to Smoking!</h4>
    <p>Our analysis reveals that smoking has the highest odds for
contributing to medical expenses, with a remarkable impact on healthcare
costs. The findings emphasize the detrimental effects of smoking on both
health and financial well-being. It serves as a stark reminder that smoking is
not only injurious to health but also exerts a significant financial burden on
individuals and the healthcare system.</p>
</div>
<!-- Analysis section End -->
<!-- ===== Frequently Asked Questions Section ===== -->
<section id="faq" class="faq section-bg">
    <div class="container" data-aos="fade-up">

        <div class="section-title">
            <h2>Frequently Asked Questions</h2>
            <p>Browse through our frequently asked questions (FAQ) section to find
quick answers to common inquiries about our project. Gain insights into cost
estimation, data sources, customization options, accuracy of predictions, and
more. Discover the information you need to make informed decisions about
healthcare expenses.</p>
        </div>

        <ul class="faq-list">

            <li>
                <div data-bs-toggle="collapse" class="collapsed question"
href="#faq1">How can this project help me estimate and predict my personal
medical care costs? <i class="bi bi-chevron-down icon-show"></i><i class="bi
bi-chevron-up icon-close"></i></div>
                <div id="faq1" class="collapse" data-bs-parent=".faq-list">
                    <p>
                        Our project utilizes advanced data analytics techniques to analyze
various factors that influence medical care costs, such as age, BMI, smoking
habits, and more. By comparing your relevant information with our visuals,
you can receive approximate cost estimates and predictions.
                    </p>
                </div>
            </li>

```


<div data-bs-toggle="collapse" href="#faq2" class="collapsed question">What data sources are used in the estimation and prediction of hospitalization and medical care costs? <i class="bi bi-chevron-down icon-show"></i><i class="bi bi-chevron-up icon-close"></i></div>

<div id="faq2" class="collapse" data-bs-parent=".faq-list">

<p>

In our project, we utilize the comprehensive insurance dataset available here. This dataset contains valuable information such as age, gender, BMI, smoking habits, region, and medical charges. By leveraging this dataset, we derived meaningful insights and accurate predictions regarding healthcare costs.

</p>

</div>

<div data-bs-toggle="collapse" href="#faq3" class="collapsed question">Are the cost predictions provided by this project specific to a particular region or applicable globally? <i class="bi bi-chevron-down icon-show"></i><i class="bi bi-chevron-up icon-close"></i></div>

<div id="faq3" class="collapse" data-bs-parent=".faq-list">

<p>

The cost predictions generated from our project can be customized based on the available data and can be applied to different regions or specific locations. However, it's important to consider that local healthcare systems, insurance coverage, and other regional factors may influence the accuracy of these predictions.

</p>

</div>

<div data-bs-toggle="collapse" href="#faq4" class="collapsed question">Can I customize the variables and parameters used in the cost estimation visualizations?<i class="bi bi-chevron-down icon-show"></i><i class="bi bi-chevron-up icon-close"></i></div>

<div id="faq4" class="collapse" data-bs-parent=".faq-list">

<p>

No, unfortunately our project doesn't offer flexibility in customizing the variables and parameters used in the cost estimation visuals. You can

however alter the filters based upon specific factors like age, BMI and smoking habits.

</p>

</div>

<div data-bs-toggle="collapse" href="#faq5" class="collapsed question">Are the cost predictions provided by this project accurate? <i class="bi bi-chevron-down icon-show"></i><i class="bi bi-chevron-up icon-close"></i></div>

<div id="faq5" class="collapse" data-bs-parent=".faq-list">

<p>

The accuracy of the cost predictions in this project is subjective and depends on the viewer's capability to interpret and analyze the provided data visualizations. Our project presents the available medical cost data accurately, allowing users to draw their own conclusions and insights based on the displayed information. The accuracy of any predictions or inferences made based on the visualizations may vary depending on individual interpretation and analysis skills.

</p>

</div>

<div data-bs-toggle="collapse" href="#faq6" class="collapsed question">What is the real world impact of your project? <i class="bi bi-chevron-down icon-show"></i><i class="bi bi-chevron-up icon-close"></i></div>

<div id="faq6" class="collapse" data-bs-parent=".faq-list">

<p>

Our project has a real-world impact by empowering stakeholders with valuable insights into medical care costs, enabling informed decision-making, resource optimization, and improved financial management. Ultimately, our project aims to contribute to a more transparent, efficient, and sustainable healthcare ecosystem.

</p>

</div>

</div>

```
</section><!-- End Frequently Asked Questions Section -->
```

```
<!-- ===== Contact Section ===== -->
```

```
</main><!-- End #main -->
```

```
<!-- Vendor JS Files -->
```

```
<script
```

```
src="static/assets/vendor/purecounter/purecounter_vanilla.js"></script>
```

```
<script src="static/assets/vendor/aos/aos.js"></script>
```

```
<script
```

```
src="static/assets/vendor/bootstrap/js/bootstrap.bundle.min.js"></script>
```

```
<script src="static/assets/vendor/glightbox/js/glightbox.min.js"></script>
```

```
<script src="static/assets/vendor/swiper/swiper-bundle.min.js"></script>
```

```
<script src="static/assets/vendor/php-email-form/validate.js"></script>
```

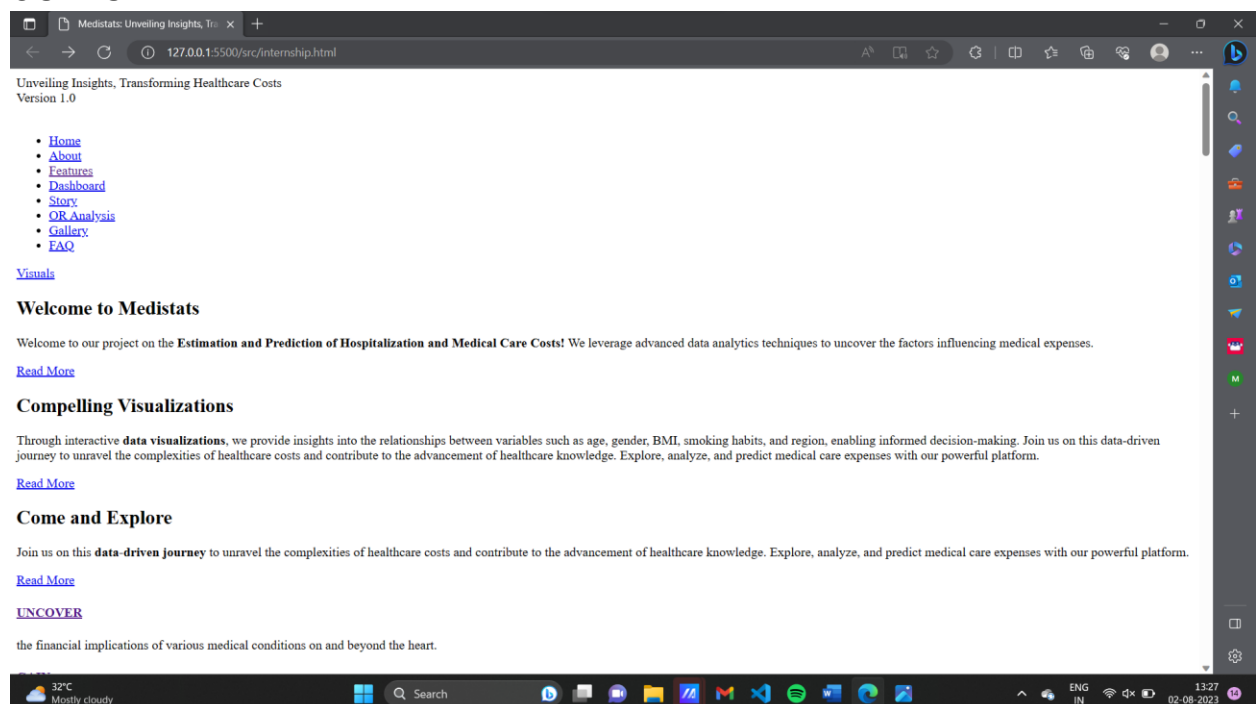
```
<!-- Template Main JS File -->
```

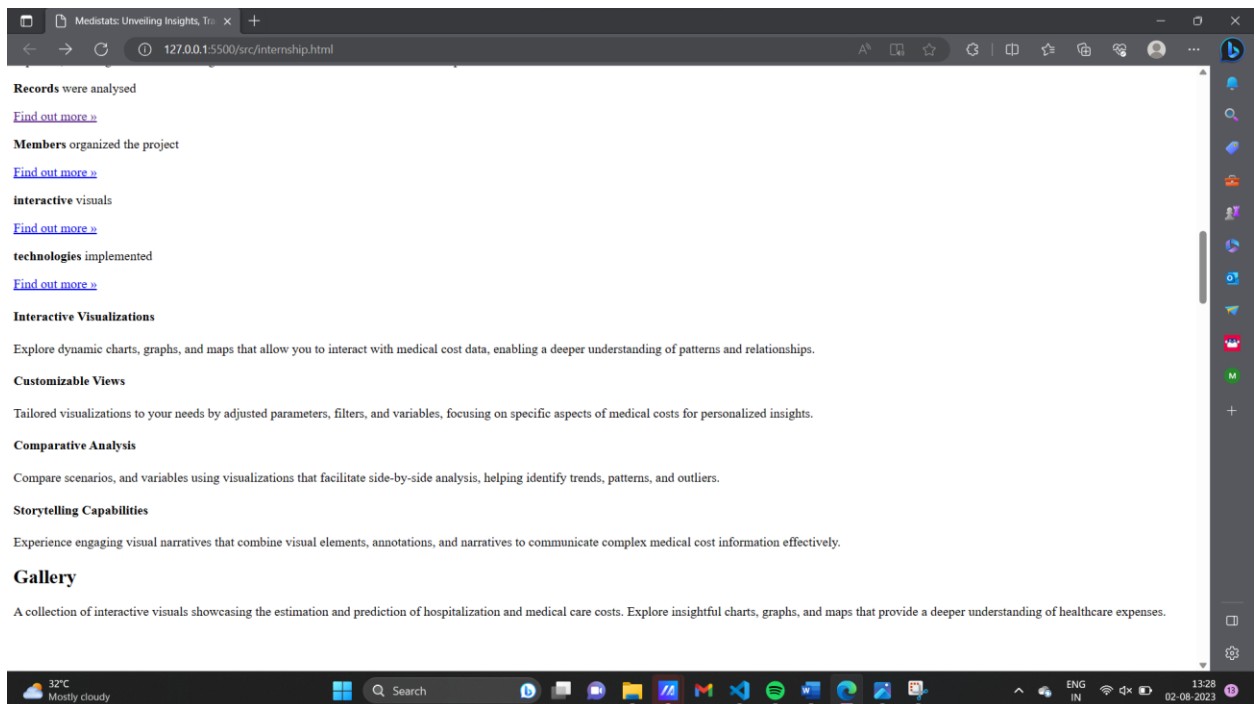
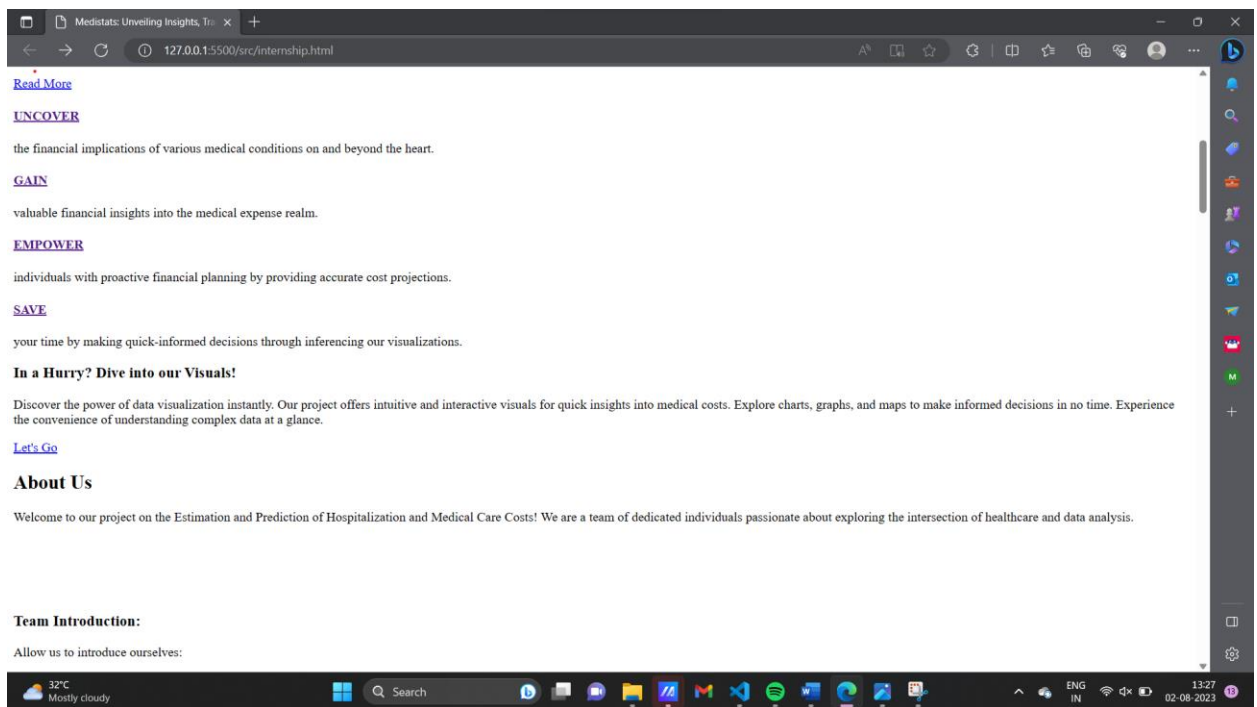
```
<script src="static/assets/js/main.js"></script>
```

```
</body>
```

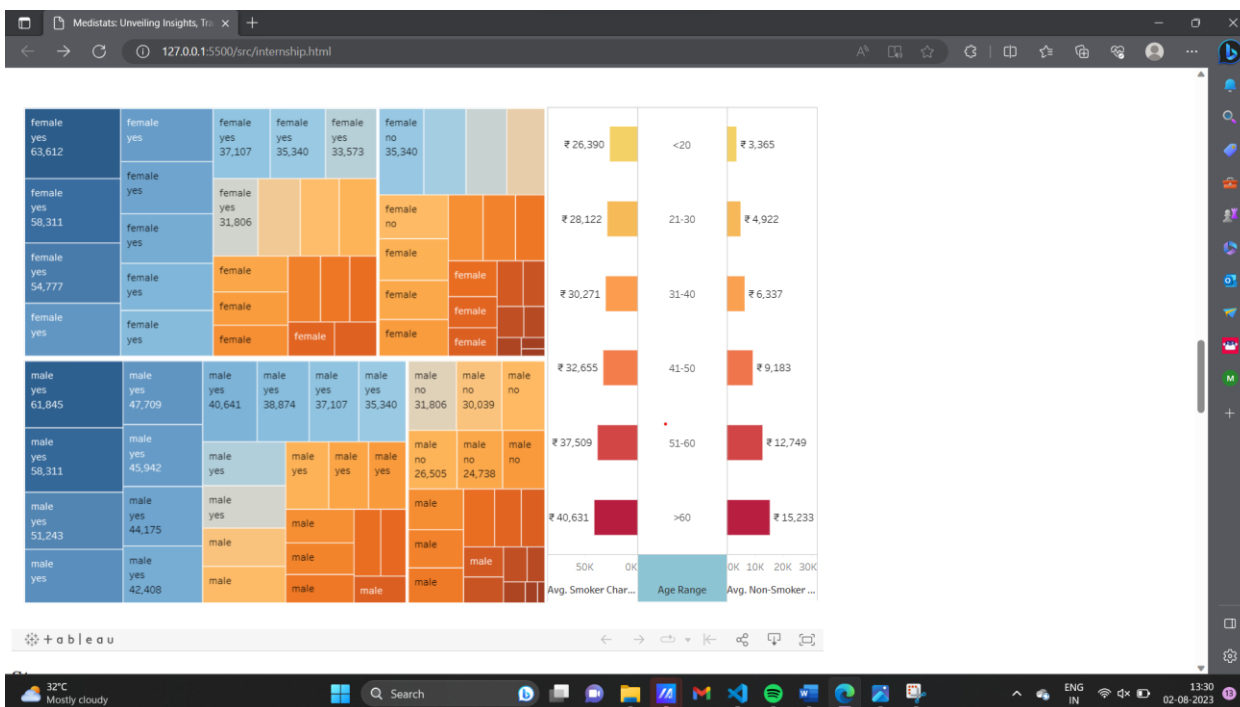
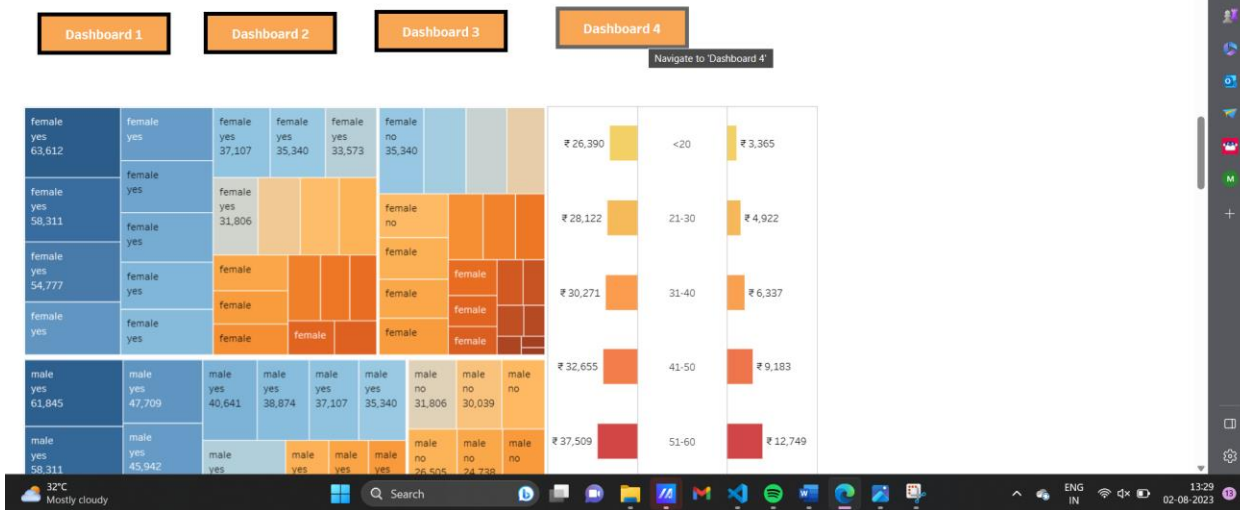
```
</html>
```

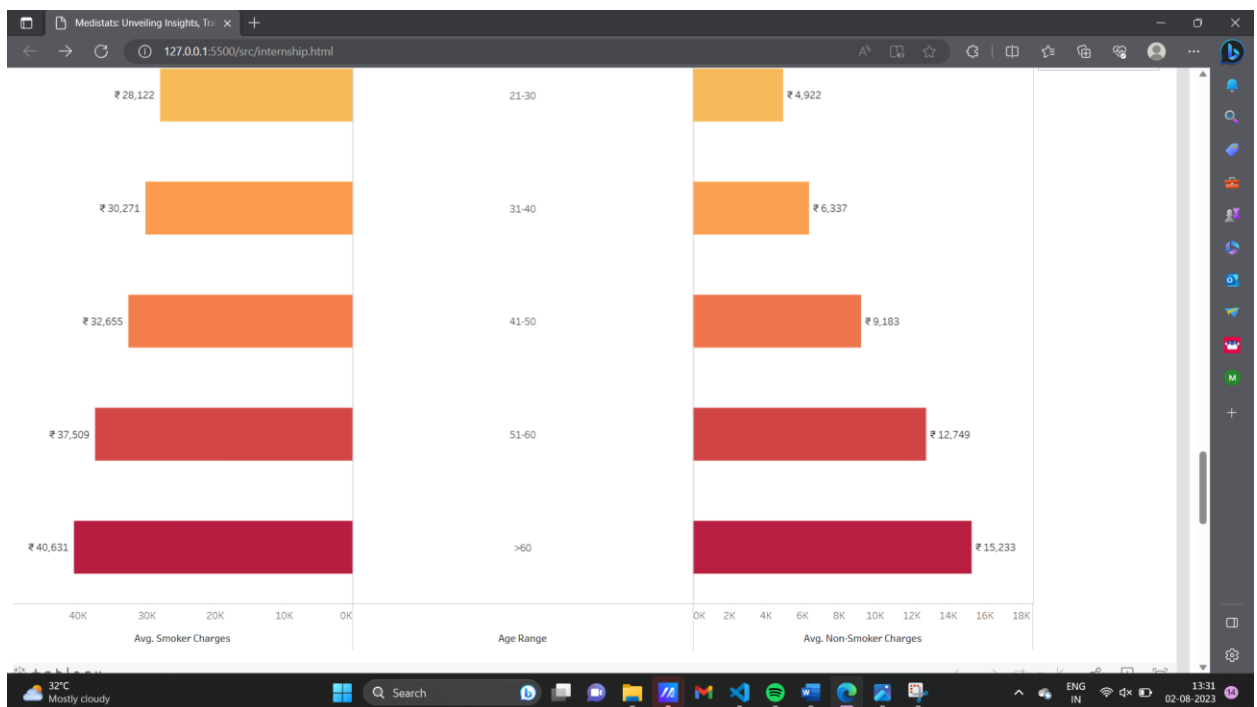
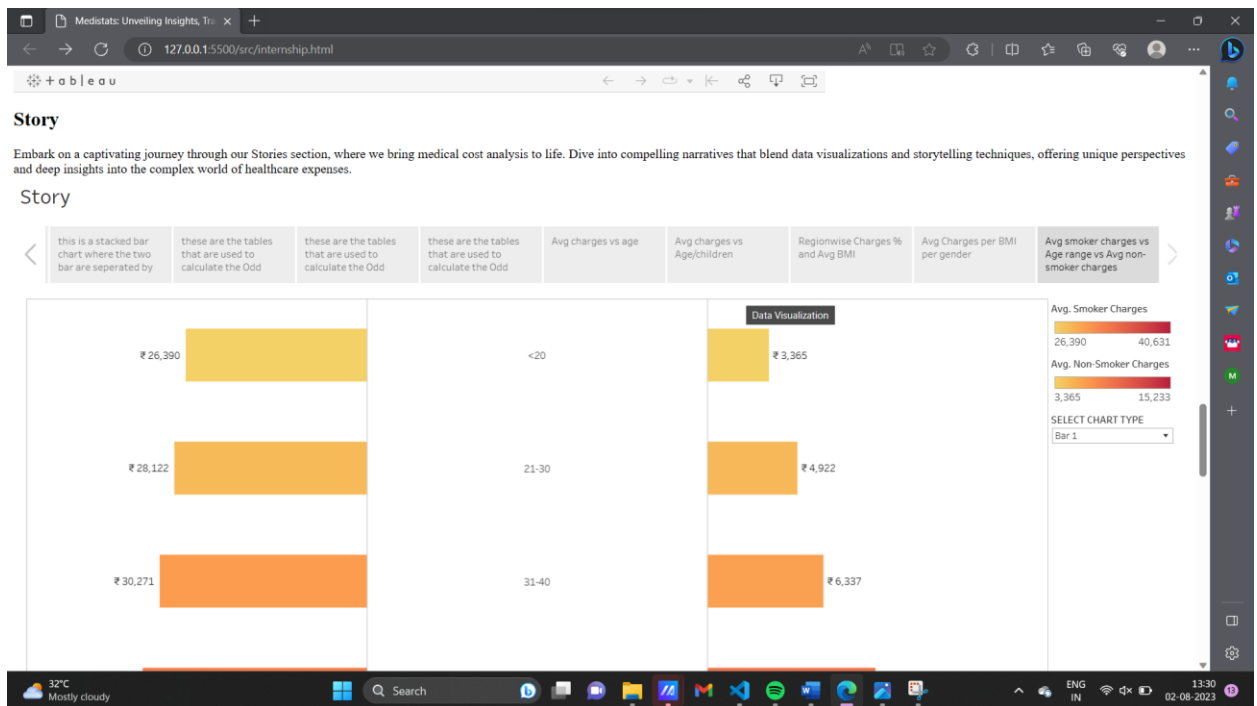
OUTPUT:

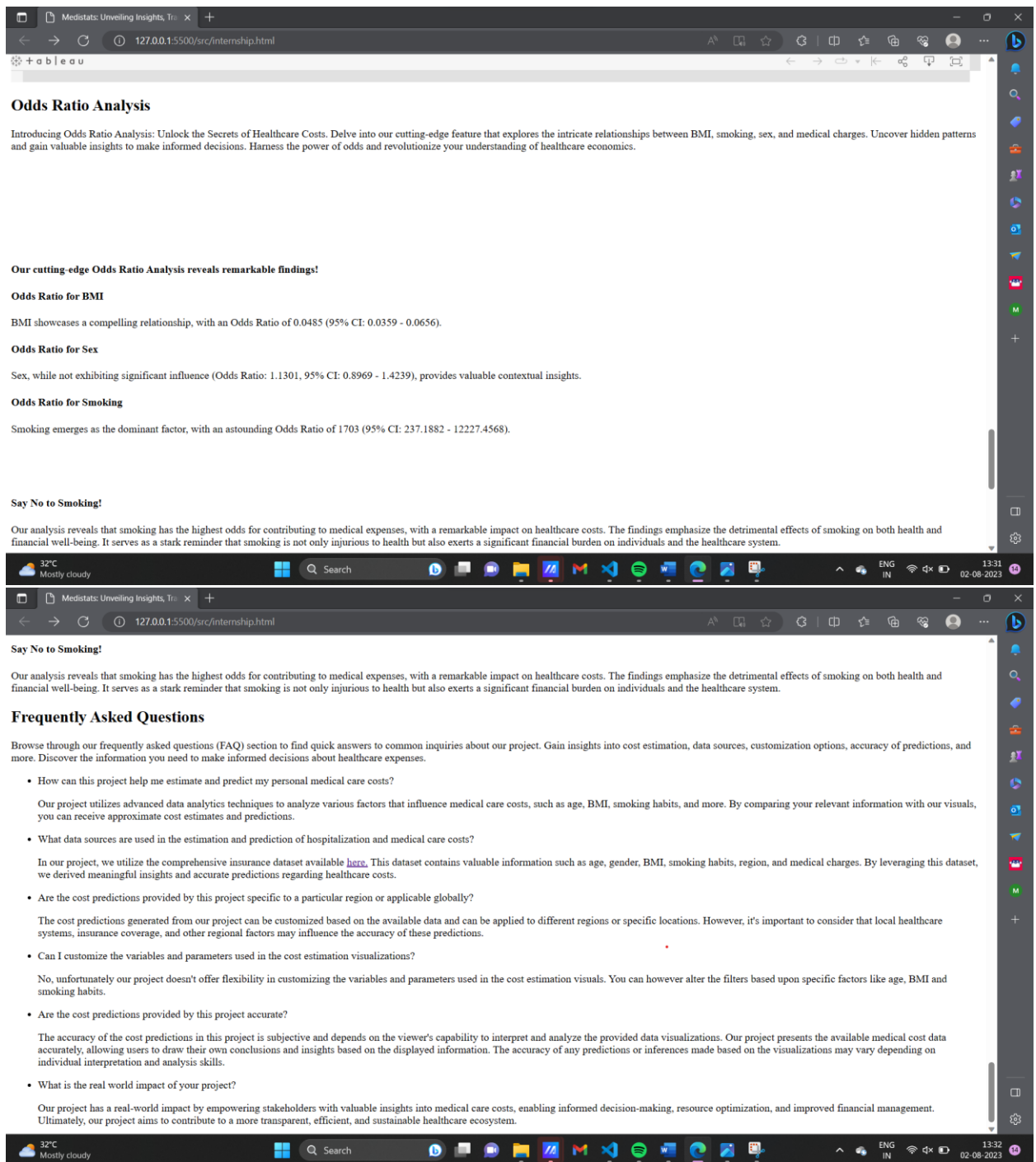




Explore our comprehensive dashboards for insightful analysis of hospitalization and medical care costs. Customize and filter data, uncover trends, and make data-driven decisions. Experience the power of visualizations in driving informed strategies and improving healthcare outcomes.







Acknowledgments:

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi

REFERENCES:

- A. K. Dutta, N. Ali Aljarallah, T. Abirami et al., "Optimal deep-learning-enabled intelligent decision support system for SARS-CoV-2 classification," *Journal of Healthcare Engineering*, vol. 2022, Article ID 4130674, 14 pages, 2022.
- [1] J. Mohana, B. Yakkala, S. Vimalnath et al., "Application of internet of things on the healthcare field using convolutional neural network processing," *Journal of Healthcare Engineering*, vol. 2022, Article ID 1892123, 2022.
- [2] L. Hu, L. Li, J. Ji, and M. Sanderson, "Identifying and understanding determinants of high healthcare costs for breast cancer: a quantile regression machine learning approach," *BMC Health Services Research*, vol. 20, no. 1, pp. 1066–1110, 2020.
- [3] M. A. Aefa, M. Mahmoud, and M. M. Nassar, "Parameter estimation for a mixture of inverse chen and inverse compound Rayleigh distribution based on type-I hybrid censoring scheme," *Journal of Statistics Applications & Probability*, vol. 10, no. 3, pp. 647–663, 2021.
- [4] W. A. Afifi and A. H. El-Bagoury, "Optimal multiplicative generalized linear search plan for a discrete randomly located target," *Information Sciences Letters*, vol. 10, no. 1, pp. 153–158, 2021.
- [5] R. A. Ganaie, V. Rajagopalan, and S. Aldulaimi, "The weighted power shanker distribution with characterizations and applications of real-life time data," *Journal of Statistics Applications & Probability*, vol. 10, no. 1, pp. 245–265, 2021.
- [6] M. H. Abu-Moussa, A. M. Abd-Elfattah, and E. H. Hafez, "Estimation of stress-strength parameter for Rayleigh distribution based on progressive type-II censoring," *Information Sciences Letters*, vol. 10, no. 1, pp. 101–110, 2021.
- [7] S. Sana and M. Faizan, "Bayesian estimation using lindley's approximation and prediction of generalized exponential distribution based on lower record values," *Journal of Statistics Applications & Probability*, vol. 10, no. 1, pp. 61–75, 2021.
- [8] K. Sahu and R. K. Srivastava, "Needs and importance of reliability prediction: an industrial perspective," *Information Sciences Letters*, vol. 9, no. 1, pp. 33–37, 2020.
- [9] A. A. Soliman, Al-W. A. Farghal, and G.A. Abd-Elmougod, "Statistical inference under copula approach of accelerated dependent generalized inverted exponential failure time with progressive hybrid censoring scheme," *Applied Mathematics & Information Sciences*, vol. 15, no. 6, pp. 687–699, 2021.
- [10] S. Kent, J Green, G Reeves et al., "Hospital costs in relation to body-mass index in 1.1 million women in England: a prospective cohort study," *The Lancet Public Health*, vol. 2, no. 5, pp. e214–e222, 2017.
- [11] V. S. Kadam, S. Kanhere, and S. Mahindrakar, "Regression techniques in machine learning & applications: a review," *International Journal for Research in Applied Science and Engineering Technology*, vol. 8, no. 10, pp. 826–830, 2020.

- [12] B. Panay, N. Baloian, J. Pino, S. Peñafiel, H. Sanson, and N. Bersano, "Predicting health care costs using evidence regression," *Proceedings*, vol. 31, p. 74, 2019, <https://www.mdpi.com/2504-3900/31/1/74>.
- [13] B. J. Moore, S. White, R. Washington, N. Coenen, and A. Elixhauser, "Identifying increased risk of readmission and in-hospital mortality using hospital administrative data," *Medical Care*, vol. 55, no. 7, pp. 698–705, 2017.
- [14] R. S. Suidan, W. He, C. C. Sun et al., "Impact of body mass index and operative approach on surgical morbidity and costs in women with endometrial carcinoma and hyperplasia," *Gynecologic Oncology*, vol. 145, no. 1, pp. 55–60, 2017.
- [15] H. J. Kan, H. Kharrazi, H.-Y. Chang, D. Bodycombe, K. Lemke, and J. P. Weiner, "Exploring the use of machine learning for risk adjustment: a comparison of standard and penalized linear regression models in predicting health care costs in older adults," *PloS one*, vol. 14, no. 3, Article ID e0213258, 2019.
- [16] S. Kent, S. A. Jebb, A. Gray et al., "Body mass index and use and costs of primary care services among women aged 55-79 years in England: a cohort and linked data study," *International Journal of Obesity*, vol. 43, no. 9, pp. 1839–1848, 2019.
- [17] J. A. Irvin, A. A Kondrich, M Ko et al., "Incorporating machine learning and social determinants of health indicators into prospective risk adjustment for health plan payments," *BMC Public Health*, vol. 20, no. 1, pp. 608–610, 2020.
- [18] S. Kent, F. Fusco, A. Gray, S. A. Jebb, B. J. Cairns, and B. Mihaylova, "Body mass index and healthcare costs: a systematic literature review of individual participant data studies," *Obesity Reviews*, vol. 18, no. 8, pp. 869–879, 2017.
- [19] A. Ashfaq, A. Sant'Anna, M. Lingman, and S. Nowaczyk, "Readmission prediction using deep learning on electronic health records," *Journal of Biomedical Informatics*, vol. 97, Article ID 103256, 2019.
- [20] R. J. Desai, S. V. Wang, M. Vaduganathan, T. Evers, and S. Schneeweiss, "Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes," *JAMA Network Open*, vol. 3, no. 1, Article ID e1918962, 2020.
- [21] C.-Y. Kuo, L.-C. Yu, H.-C. Chen, and C.-L. Chan, "Comparison of models for the prediction of medical costs of spinal fusion in Taiwan diagnosis-related groups by machine learning algorithms," *Healthcare informatics research*, vol. 24, no. 1, pp. 29–37, 2018.
- [22] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann, Massachusetts, MA, USA, Fourth edition, 2016.
- [23] M. Barrett, J. Boyne, J. Brandts et al., "Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care," *The EPMA Journal*, vol.10, no. 4, pp. 445–464, 2019.
- [24] J. Liu, D. Capurro, A. Nguyen, and K. Verspoor, "Early prediction of diagnostic-related groups and estimation of hospital cost by processing clinical notes," *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1–8, 2021.

- [25] H. N. Alhazmi, A. Alghamdi, F. Alajlani, S. Abuayied, and F. M. Aldosari, "Care cost prediction model for orphanage organizations in Saudi Arabia," *IJCSNS*, vol. 21, no. 4, p. 84, 2021.
- [26] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," in *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 492–499, Madurai, India, June 2017.
- [27] Z. J. Ward, S. N. Bleich, M. W. Long, and S. L. Gortmaker, "Association of body mass index with health care expenditures in the United States by age and sex," *PloS one*, vol. 16, no. 3, Article ID e0247307, 2021.
- [28] I. Osawa, T. Goto, Y. Yamamoto, and Y. Tsugawa, "Machinelearning-based prediction models for high-need high-cost patients using nationwide clinical and claims data," *NPJ digital medicine*, vol. 3, no. 1, pp. 148-149, 2020.
- [29] C. Yang, C. Delcher, E. Shenkman, and S. Ranka, "Machine learning approaches for predicting high-cost high need patient expenditures in health care," *BioMedical Engineering Online*, vol. 17, no. 1, pp. 131–220, 2018.
- [30] H. Kharrazi, H.-Y. Chang, S. E. Heins, J. P. Weiner, and K. A. Gudzone, "Assessing the impact of body mass index information on the performance of risk adjustment models in predicting health care costs and utilization," *Medical Care*, vol. 56, no. 12, pp. 1042–1050, 2018.
- [31] F. Wang, T. McDonald, J. Bender, B. Reffitt, A. Miller, and D. W. Edington, "Association of healthcare costs with per unit body mass index increase," *Journal of Occupational and Environmental Medicine*, vol. 48, no. 7, pp. 668–674, 2006.
- [32] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *Journal of the American Geriatrics Society*, vol. 66, no. 4, pp. 693–698, 2018.
- [33] M. Ravaut, H. Sadeghi, K. K. Leung et al., "Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–12, 2021.
- [34] C. Wu, F. Wu, Y. Huang, and X. Xie, "NICE: neural inhospital cost estimation from medical records," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2409–2412, Beijing, China, 2019.
- [35] N. I. Jha, I. Ghergulescu, and A.-N. Moldovan, "OULAD MOOC dropout and result prediction using ensemble, deep learning and regression techniques," in *Proceedings of the 11th International Conference on Computer Supported Education CSEDU*, no. 2, pp. 154–164, Heraklion, Crete, Greece, MAY 2019.
- [36] M. P. Shyamala Devi, M. Swathi, V. Purushotham Reddy et al., "Linear and ensembling regressionbased health cost insurance prediction using machine learning," in *In: Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies*, S. C. Satapathy, V. Bhateja, M. N. Favorskaya, and T. Adilakshmi, Eds., vol. 224, Singapore, Springer, 2019.
- [37] S. Baharvand, A. Jozaghi, R. Fatahi-Alkouhi, S. Karimzadeh, R. Nasiri, and B. Lashkar-Ara, "Comparative study on the machine learning and regression-based approaches to predict the hydraulic jump sequent depth ratio," *Iranian Journal*

of Science and Technology, Transactions of Civil Engineering, vol. 45, pp. 2719–2732, 2021.

- [38] N. M. Mohamed, “Estimation on kumaraswamy-inverse weibull distribution with constant stress partially accelerated life tests,” *Applied Mathematics & Information Sciences*, vol. 15, no. 4, pp. 503–510, 2021.
- [39] G. Manogaran and D. Lopez, “Health data analytics using scalable logistic regression with stochastic gradient descent,” *International Journal of Advanced Intelligence Paradigms*, vol. 10, no. 1–2, pp. 118–132, 2018.